FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

DIGITAL FARMER PROFILES:
Reimagining Smallholder Agriculture

USAID
FROM THE AMERICAN PEOPLE

AUTHORS

Bobbi Gray, Grameen Foundation
Lee Babcock, Grameen Foundation
Leo Tobias, Grameen Foundation
Mona McCord, Grameen Foundation
Ana Herrera, Grameen Foundation
Cecil Osei, Grameen Foundation
Ramiro Cadavid, Grameen Foundation

ACKNOWLEDGEMENTS

Digital Development for Feed the Future is a collaboration between the Global Development Lab and the Bureau for Food Security, both within the United States Agency for International Development (USAID), and is focused on integrating a suite of coordinated digital tools and technologies into Feed the Future activities to accelerate agriculture-led economic growth and improved nutrition. Feed the Future is America's initiative to combat global hunger and poverty.

Many thanks go to all the service providers and others (listed in Annex 1) who gave of their time to contribute to this landscape assessment through interviews and email exchanges and for directing our team to additional resources. We hope we have adequately presented their work, experiences, and opinions.

The authors of this publication would like to express their gratitude to the many Grameen Foundation team members who assisted in conducting interviews and identifying resources for this research: Maria Hernandez, Gaurav Chakraverty, Simon Okot, Evelyne Banura, Brigitta Nyawira, and Benjamin Kimosop. Also, of the Grameen Foundation, we would like to express our gratitude to Lauren Hendricks, Sybil Chidiac, Jessie Tientcheu, Gigi Gatti, and Emily Romero for their thought leadership and support, and to Brent Farrar, Bee Wuethrich, and Liselle Yorke for their graphic design and communications assistance. Additional thanks go to Bee Wuethrich of Grameen Foundation for her copyediting.

Finally, we would like to voice our appreciation to Ellen Galdava and Abdul Bari Farahi from FHI 360 and Kwasi Donkor and Christopher Burns of USAID for their support, review, and input into this publication.

November, 2018

TABLE OF CONTENTS

Photo by USAID NEAT

ABBREVIATIONS

AGRA	Alliance for a Green Revolution in Africa
API	Application Programming Interface
ARET	Agriculture Risk Evaluation Tool
B2B	Business to Business
B2C	Business to Customer
BI	Business Intelligence
CBA	Commercial Bank of Africa
CGAP	Consultative Group to Assist the Poor
CGIAR	Consultative Group for International Agricultural Research
CIAT	International Center for Tropical Agriculture (part of CGIAR network)
CoP	Community of Practice
CTA	The Technical Centre for Agricultural and Rural Cooperation
FAIR	Findable, Accessible, Interoperable, Reusable
FAO	Food and Agriculture Organization
FI	Financial Institution
FUNDER	Foundation for Rural Business Development
GCAP	Ghana Commercial Agricultural Project
GPS	Global Positioning System
GSMA	Groupe Speciale Mobile Association
HCD	Human-centered Design
i2i	Impact 2 Insight
IARI	Indian Agriculture Research Institute
IVR	Interactive Voice Response
KYC	Know Your Customer
LST	Land Surface Temperature
LSMS	Living Standards Measurement Study
MNO	Mobile Network Operator
NDVI	Normalized Difference Vegetation Index
NGO	Non-governmental Organization
OBD	Outbound Dialing
SMS	Short Message Service
VGI	Volunteered Geographic Information
WAAPP	West Africa Agricultural Productivity Program

GLOSSARY

Algorithm	A set of rules to be followed in calculations or other problem-solving operations, especially by a computer.
Artificial Intelligence	Emerged in the 1950s. The theory and development of computer systems to be able to perform tasks normally requiring human intelligence.
Big Data*	Open, harmonized, interoperable, and integrated datasets from multiple domains aimed to accelerate agricultural research and data use in service of development goal.
Blockchain	A secure distributed immutable database shared by all parties in a distributed network where transaction data can be recorded.
Business Intelligence	Using data generated by service users to make decisions about product/service design.
Data-driven Agriculture	Thoughtful use of data (often big data) to inform farmer decisions and actions. It means having the right data, at the right time, to make better decisions that improve long-term profitability.
Deep Learning	Emerged in 2010s; a technique for implementing machine learning.
Farmer Ecosystem	An interconnected and coordinated network of support services, information, suppliers, buyers and actors that meet the needs of farming households.
Farmer Profile	Data collected on a farmer and his or her farm that is used by a service provider or multiple service providers to design and direct products or services.
Internet of Things	The interconnection via the internet of computing devices embedded in everyday objects, enabling them to send and receive data; e.g., in soil, farm tools and waterways.
Land Surface Temperature	The temperature of the land itself rather than the ambient air above it (as is used in most typical temperature recording).
Machine Learning	An algorithm that is trained, given input data, and then run on new data to predict the output. As the system processes more data, it learns from its mistakes. Emerged in 1980s. An approach for achieving artificial intelligence and is used in predictive, prescriptive, and cognitive analytics.
Normalized Difference Vegetation Index	A measure of plant health derived from satellite imagery.
Ontology	A set of concepts and categories in a subject area or domain that shows their properties and the relations between them. Currently perceived as being necessary for interoperability of data.
Outbound Dialing	Also called voice SMS, is a pre-recorded message sent to mobile phones.
Precision Agriculture	An integrated crop management system that attempts to match the kind and amount of inputs with the actual crop needs for small areas within a farm field. Precision agriculture uses one or more of the following sources of data: soils, crops, nutrients, pests, moisture, or yield, for optimum profitability, sustainability, and protection of the environment. Also referred to as satellite farming.
Short Message Service	Known as SMS, written messages with limited character length sent to phones.

Definitions are drawn from GSMA, i2i, CIAT/CGIAR, dictionary.com, University of Missouri, USDA, and others.

**Big data has several different definitions that are outlined in this report. For this report, the CGIAR definition will be used.*

EXECUTIVE SUMMARY

More than 500 million smallholder farms worldwide play a significant role in food production and the genetic diversity of the food supply. Until now, it has been difficult to get information to or from smallholder farmers, compounding basic infrastructural problems such as access to inputs, markets, financing, and training. The spread of mobile technology, remote-sensing data, and distributed computing and storage capabilities are opening new opportunities to integrate smallholder farmers into the broader agri-food system. The scale of these changes holds out the potential for another agricultural revolution.

MORE THAN **500** MILLION smallholder farms worldwide play a significant role in food production and the genetic diversity of the food supply.

As mobile technology use increases and improves in rural areas, the paradigm is also shifting for how smallholder farmers are profiled, how their needs are understood and met, how the impact of agricultural services is measured, how farmer data is shared, and how a global body of knowledge can be built by drawing on typically siloed expertise and data. To help describe this shift in farmer profile data management, Grameen Foundation conducted a landscape assessment that:

- Documents experiences in managing digital farmer data by describing how smallholder farmers are defined, the types of service providers that collect farmer profile data, how data is collected, analyzed, and used to support smallholders with products and services, and how this data is shared and managed.
- Highlights innovative models of smallholder farmer data management and sharing to inspire new thinking among actors in this space.
- Outlines key considerations when assessing existing or investing in new efforts to develop and leverage farmer profile data.

Over a three-month period, Grameen Foundation conducted a desk review of the existing literature, including peer-reviewed journal articles, project and program reports and presentations, blogs, and information provided on websites. Grameen Foundation also conducted approximately 50 key informant interviews with service providers,[1] including mobile network operators (MNOs), agribusinesses, government, research firms, technical assistance providers, donors, and other actors. These service providers were assumed to provide a valid representation of current farmer data management practice. Grameen Foundation also participated in the ICTforAG 2018 and Data Driven Agriculture: The Future of Smallholder Farmer Data Management and Use workshop in summer 2018, the latter organized by USAID funded and FHI 360 lead Mobile Solutions, Technical Assistance and Research (mSTAR) project, to gather additional input from technical specialists in the area. This landscape assessment was meant to capture different contexts and stakeholders but not to be exhaustive. The insights collected are meant to guide development organizations, USAID, and other donors in using these models in their everyday operations or in project design.

1 A service provider can be any entity involved in data capture, packaging, and analysis to inform farmer decision making—inter alia an MNO, agribusiness, government, NGO, and research firm.

Understanding smallholder farmer data management requires defining from whom data is captured and how it is captured, analyzed, used, and shared. The findings from the assessment revealed:

- Defining a smallholder farmer is not easy; farmers are not homogeneous. The definition of a smallholder farmer must be flexible to include particularly vulnerable people.
- Given there can be several farmers per plot of land, when aggregating data for either open data efforts or for sophisticated analytics, it is the farm itself that "pulls" data together. The farm is the common denominator, not the farmer, for aggregating data.
- Service providers who capture data from, about, and for farmers are a diverse group. The type of service provider involved does not necessarily determine what data is collected or how it is used, but it is an important starting point.
- New ways of collecting and aggregating data and applying analytics—such as predictive, prescriptive, and cognitive analytics—can reduce the amount of direct input needed from the farmer. Data analytics is a game-changer and is being used to create "new data" from existing data.
- Digital technology is now facilitating the sharing and management of farmer profile data in real time.
- Marrying plant science data with real-time farmer data is a new frontier for improving farm productivity.
- Service providers win or lose depending on how they use their data. Service providers and farmers should not treat data as just a resource but as an asset and should consider opportunities to monetize data. Smallholder farmer data is giving rise to new configurations within service provider business models.
- There is no single pathway to sophisticated use of farmer profile data. How service providers use and manage data in combination with the data capture methods and data analytics determines how innovative their model may be.
- When service providers consider new farmer profile approaches, they should start small and take manageable steps.

This assessment has also revealed that most of the data and the technology (hardware and software) already exist to solve many constraints that farmers face, but the solutions are fragmented and not all service providers—or farmers—have equal opportunities to access them. Big data promises to bring fragmented data, resources, and service providers together to support the farmer ecosystem. There is no better time than now to strengthen farmer ecosystems. Advancements in technology can facilitate the sector's ability to:

- Know what farmer data service providers do and do not have.
- Marry plant/animal science with human science.
- Fully link the farmer with the ecosystem.
- Be precise with information and financial decisions.

All the pieces are there for supporting data-driven agriculture. The challenge is to bring the pieces together for the benefit of the farmer and the world he or she is expected to feed.

I. INTRODUCTION

The International Labor Organization estimates that slightly over 1 billion people globally are engaged in agriculture ("Agriculture," 2018), accounting for approximately 29 percent of the workforce ("Employment," 2017). This number is declining even though the demand for agricultural crops is expected to double as the world population reaches 9.1 billion by 2050 (Fischer, 2016). There are an estimated 570 million smallholder farms worldwide (Lowder, Skoet, & Raney, 2016); they play a significant role in food production as well as in maintaining the genetic diversity of the food supply, mitigating risks of nutritional deficiencies and ecosystem degradation (Fanzo, 2017).

Smallholder farmers are facing an ever-changing world: the seeds that worked for generations may not be the seeds that work today due to climate change, soil degradation, and water constraints. A farmer can no longer rely on historical calendars and generational knowledge to drive decisions about purchases, seeds to plant, and mulches and fertilizers to use. He or she needs more timely and responsive support.

McKinsey & Company (2015) found that even in industrialized countries like the United States, the agricultural sector ranks at the bottom in terms of digitalization. Progress in low- and middle-income countries is assumed to be even slower. But this is starting to change. In October 2017, GSMA estimated there were slightly more than 5 billion unique mobile subscribers worldwide ("Representing," 2017). They estimated that by 2020, almost three-quarters of the world's population—5.7 billion people—would subscribe to mobile services (GSMA, 2017). The next wave of mobile connections is expected to come mainly from rural areas (Palmer & Darabian, 2017b) where those engaged in agriculture live. The ubiquity of mobile phones has the potential to change the face of farming in Africa, where smallholder formers are responsible for more than 80 percent of agricultural production.

The hopes and expectations for digitalization of agriculture are high. Using mobile tools, data can be directly collected from smallholder farmers using their own mobile tools or indirectly by service providers using mobile tools to record data provided by farmers. It can then flow directly into a database. Data management processes can exceed the capacities of typical processes in the past. Data is being used to create robust farmer profiles that can be refined over time but accessed in real-time by multiple service providers—such as financial services providers, input suppliers, agro-processors, and farmer cooperatives—to understand and engage with farmers better. Input suppliers can be more prescriptive if they have data available on soil and crop health; agro-processors can use data to estimate upcoming harvest volume and even manage traceability of agricultural products. Financial services providers can use the data for farmer risk-profiling and can become very efficient and logical, using data on likely crop output. Farmers can more effectively utilize fertilizers and other soil amendments, thus reducing expenses, producing higher yields, and creating more environmentally friendly farms. All data collected from farmers through mobile tools can be made

AGRICULTURE SNAPSHOT

OVER 1 BILLION *people globally are engaged in agriculture*

Demand for agriculture crops is expected to **DOUBLE** *by* **2050**

The ubiquity of **MOBILE PHONES** *has the potential to change the face of farming in Africa where more than* **80%** *of agricultural production comes from smallholder farmers.*

Photo by Daljit Singh, Feed the Future

available via an ICT (information and communications technology) platform that can provide timely data to all platform users, including the farmer. Farmers will also find the solutions attractive and therefore demand those services, contributing to a sustainable offer of products and services over the long term.

Mobile technology can provide a way forward to mitigate risks, strengthen value chains, and coordinate value chain actors. However, a sustainability path for many technology service models has yet to be established, and the coordination of disparate data sets has yet to be fully achieved.

This publication was developed for service providers who support smallholder farmers and utilizes an adapted Evaluation Plan framework (Evaluation, n.d.) to document how smallholder farmers globally are defined, who collects data from or on them (service providers), and how data is collected or captured, stored, analyzed, shared, and used. Examples of innovative practices in these areas by service providers are sprinkled throughout. Three in-depth cases follow the main body of the report and provide a more comprehensive view of different service provider models: research, technical assistance/project management, and commercial.

Each section begins with a summary of key points as well as a list of key considerations for service providers, to facilitate review of the content and reflection by an organization on its own practices. The questions are not exhaustive but meant to help stimulate other important questions. They are posed for all service providers—whether research organizations, non-governmental organizations (NGOs), or commercial entities. Donors can also use these questions to stimulate reflection on existing or future portfolios of grantees who collect farmer profile data. Annex 2 provides a comprehensive list of all summary points and key considerations referenced in the different sections.

II. DEFINING SMALLHOLDER FARMERS

Defining smallholders is not a simple task.

To explain how farmer profile data is managed, "smallholder farmer" must be defined. But this is not a simple task (Chamberlain, 2007; Kalita, M'Cormack, & Heirman, 2012). The way smallholder farmers are defined and targeted for projects or products can have significant impact on farmer outcomes (Phillips, Waddington, & White, 2014).

Defining Smallholder Farmers

The list below provides examples of how smallholder farmers have been defined and targeted.

Size of landholding: A very common definition used by service providers is the maximum number of hectares of land owned by a household or person for growing crops or raising livestock—often designated as less than two hectares of land (Lowder et al., 2016).

Specific thresholds: Service providers sometimes establish thresholds for land size management (such as the weighted median threshold) by calculating and ordering farm sizes from largest to smallest and designating the middle size as the upper acceptable threshold (Squarcina, 2017). Consultative Group to Assist the Poor (CGAP), conducted research on smallholder farmer diaries (Anderson, Marita, & Musiime, 2016) using a matrix of characteristics to help define how smallholders could differ from other households across the six countries where they conducted research. These characteristics included market orientation, landholding size, labor input, income, farming system, farming management responsibility, capacity, legal aspects, and level of organization. They chose landholding size and livestock count as the starting points for selecting their sample target group, in addition to the landholder's self-perception of agriculture as important to his or her household's livelihood.

Summary –Defining Smallholders

- Definitions of smallholder farmers vary significantly due to heterogeneity of the group and this has implications as to who is profiled.
- While stakeholders must often create a (limiting) definition to help with targeting, any definition must be sufficiently flexible to avoid excluding particularly vulnerable people.
- There can be several farmers per plot of land; when aggregating data, the farm itself is what "pulls" data together. The farm is the common denominator, not the farmer, for aggregating data.

Key Considerations

- How is smallholder farmer defined for your organization? How much flexibility or specificity does your organization require?
- How are smallholder farmers targeted or identified? Would this targeting or identification have implications for how well you reach your targeted beneficiary?
- What definitions or targeting criteria might you consider and what implications might this have on your strategy?
- What framework/protocol is in place, if any, to capture data about the farm itself? How can your farmer definition map/integrate with other sources of farmer data? (Later sections refer to ontology and can provide guidance.)

Identifying and Targeting Smallholder Farmers

Self-identity as a farmer: CGAP's experience (Anderson, 2018) also emphasized the importance of a person's self-identification as a farmer. Those who identify as farmers may not necessarily receive most of their income from agriculture. Self-perception can influence how people join groups or programs and ultimately influence a program's outcomes.

Type of crop grown: Other providers identify farmers by the crop they grow or manage. This approach is often used to be more inclusive of women—designating some crops as "women's crops" in order to ensure their inclusion in farming services (Phillips et al., 2014).

Membership in organized farmer groups, cooperatives, or associations: Service providers also identify farmers through their active participation in specific farmer groups (e.g., focusing on crops such as rice, cocoa, coffee, and maize).

Profit potential: This approach, suggested by the International Food Policy Research Institute (IFPRI), suggests it is time to move away from the dichotomy of "small" versus "large" farms; farm size is a dynamic concept that changes as a country's overall economy grows and as non-agricultural sectors develop (Fan, Brzeska, Keyzer, & Halsema, 2013). IFPRI distinguishes between subsistence farms *with profit potential* and those without profit potential. Those with profit potential face soft constraints such as limited capital, markets, information, infrastructure, and friendly technologies. Those *without profit potential* are those that face soft and hard constraints such as poor soil, low rainfall and high temperatures, remote locations, and high population density. Farms with profit potential would be targeted with strategies such as extension services, agriculture financing, etc. Those without profit potential need alternative strategies such as social safety net support, nutrition-focused crop production for household consumption, and education/training for non-farm employment, among others.

Plot managers: Because of the challenges of including women and other commonly excluded populations (e.g., those who lease land) in farmer-support programs, there has been a recent move to define smallholder farmers based on who manages or makes decisions about a plot—i.e., a "plot manager" (Manfre, Rubin, & Nordehn, 2017). Feed the Future has also provided guidance on how to classify who in the household is a farmer and account for female management of plots, given the various roles women can play on plots owned by men (Nelson & Swindale, 2013).

Predictive digital identification: More recently, farmers are being identified through technologies, such as through MNO data. Dialog Sri Lanka, an MNO partner of GSMA's mAgri mNutrition Initiative, used cell phone location during the day and evening to predict whether the user was a housewife. In many contexts, one man can purchase several mobile phones and provide those phones to other household members, making it difficult to reach different types of users with content since all mobile phones would be registered to a male user. Dialog Sri Lanka determined that if the mobile phone did not move much during the day, the user could be assumed to be a housewife. They used this information to target home gardening content to that mobile phone. This resulted in the inclusion of more women in a service that had been primarily reaching men. Gender information may not always be collected as part of MNO registration data (Palmer & Darabian, 2017a) despite efforts of MNOs to assist in providing digital identities (GSMA, 2016). The next wave of mobile connections is expected to come from rural areas, and MNOs see an opportunity to provide more visibility and services to rural populations—increasing the likelihood that farming households will be more directly targeted with value-added services in agriculture (Palmer & Darabian, 2017b).

When aggregating, and sharing farmer data, it is the farm, and not necessarily the smallholder farmer, that pulls it all together.

The challenge of defining the smallholder *farm* is also difficult, just as the challenge of defining the smallholder *farmer* is. There is no single accepted definition of a small or family-owned farm, and this lack has undermined any effort to estimate the number of farms worldwide (Campbell & Thornton, 2014). But with the entry of big data analytics and efforts towards building flexible data management platforms, such a definition may not be critical.

Andre Jellema, of Data-Impact, works in collaboration with the ISEAL Alliance, a global membership association for setting credible standards for various industries. In collaboration with UTZ, (now merged with and known as Rainforest Alliance), a global sustainability certification organization, Jellema has created an open-source forum for service providers to discuss the development of a shared framework for "first-mile" farm data (Jellema, 2017). He started with the question, "who is a smallholder farmer?" His effort to standardize the "who" was challenging. As he put it:

> *"I got lost in defining the farmer. Someone manages the farm. Someone owns the farm. Sometimes they are the same person. Sometimes they are not. It can be an entire group of people deciding what happens on the farm.* ***It is the farm that pulls it all together.****"*

This experience shows the difficulty of rolling up data to the farmer level. Many people are potentially associated with a plot of land. Later in this report (in discussions about data analytics and aggregating data sources), this point becomes clearer. It is the farm itself that brings together the farmer (the person's potential, as influenced by soft and hard constraints), the plants (potential of plants/plant science data), and the environment (climate, weather, pests, disease, and satellite data). This concept is particularly important for data aggregation: it is not always the farmer that serves as a connection among multiple data sources, but the farm.

*What is Big Data? **

While the concept of "big data" is almost ubiquitous, agreement on what it means is not unanimous. In 2001, the company Gartner proposed a three-fold definition for big data encompassing the "three Vs": volume, velocity, and variety. This definition aimed to describe the magnitude of data being used. Gartner later added a fourth V: veracity. Oracle's definition suggests big data is simply the inclusion of additional data sources to augment existing data. Intel defines big data as "generating a median of 300 terabytes (TB) of data weekly." Microsoft suggests big data is more about computing power for analyzing massive and complex sets of information.

Ward and Barker assert that big data is "a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce, and machine learning."

This publication uses the definition proposed by CGIAR's Platform for Big Data in Agriculture: "Open, harmonized, interoperable, and integrated datasets from multiple domains aimed to accelerate agricultural research and data use in service of a development goal" (CGIAR, 2017a). CGIAR's definition acknowledges the amount of data that currently exists that could be better leveraged through open-sourced, coordinated, and integrated efforts of actors in the agricultural sector and beyond to reduce hunger and poverty.

*Source: This box draws on the research of Jonathan Stuart Ward and Adam Barker in *Undefined by Data: A Survey of Big Data Definitions.* School of Computer Science University of St Andrews, UK. https://arxiv.org/pdf/1309.5821.pdf

III. SERVICE PROVIDER MODELS—DATA GENERATORS

This section outlines the types of service providers that use farmer profile data, or data collected on a farmer and his or her farm that is used by the service provider or multiple service providers to design and direct products or services.

Summary: Service provider models

- Service providers are diverse but fall into four main categories: research entities, government extension/NGO extension services, project management/technical assistance providers, and commercial service providers.
- The type of service provider does not necessarily determine what they collect, how, and how it is used, but it is an important starting point.
- Farmers can gain real-time benefits from interacting with and using their own data.

Key considerations

- What is your goal? What expectations does your organization have for the use of farmer profile data? Adopting a technology should not be an end in itself. "Too many organizations are making big data an IT project instead of making big data a strategic business initiative that exploits the power of data and analytics to power the organization's business models." (Schmarzo, 2017)
- What are examples of relevant uses of data within your service provider model that you can draw on for inspiration? Service providers need a "realm of the possible" with respect to integrating new ideas for using farmer profiles (Schmarzo, 2016). Which service providers are using farmer profile data and employing analytics in ways that inspire you? It is hoped that this report will provide a few examples upon which to draw.
- Consider ways in which farmers can interact with the data for greater impact. Blockchain, while not necessarily a data analytic tool, is one such technology promising to put the farmer back in the driver's seat of his or her own data.
- What are the start and end points with regards to funding support and timing?
- Do farmers currently have access to and use their own data in any way? If not, what new technologies might enable and empower farmers to effectively engage with their data?

Most data used to create smallholder farmer profiles is collected by four main types of service providers: researchers, government/NGO extension services, technical assistance/project management providers, and commercial businesses.

Governments and their partners have collected data on and about farmers for decades. More recently, agribusinesses, MNOs, financial services providers, and other private sector actors have done the same. While the specific reasons for data collection can vary, the overall purpose is typically to understand farmer needs for products, information services, market linkages, and finance.

Different service providers may have different *incentives* for collecting data and making choices about data management—including what data they collect and how they collect, manage, and share the raw data or the data products (such as dashboards or reports). As discussed in three case studies presented later, the way data is collected and then used for better farmer decision making can be similar among these stakeholders, despite differences in their modus operandi.

Table 1 below highlights service providers within the four-model construct: research, government or NGO extension services, project management, and commercial. The commercial model includes business-to-business (B2B) as well as business-to-customer (B2C) services. These are often hybrid categories. For example, CGAP and CGIAR are categorized under the research model, but they also serve project management and technical assistance (TA) roles, depending on the project.

The table provides examples of each model, referencing specific service providers included in this landscape assessment. Many other service providers not directly included in this assessment would also fall into these models. (See also Annex 1 for names of specific individuals associated with service providers mentioned below.)

Table 1: Service provider models

Research	Government Extension/NGO Extension services	Project Management/ Technical Assistance	Commercial
Data is collected for research projects, whether for impact measurement or pure research such as financial diaries research, market research, plant/animal science research, impact assessments, etc. Research can be both short-term and long-term efforts.	Data can be collected for decades to capture crop conditions and weather trends and is used to target programs and services to farmers and day-to-day management.	Data is collected for a short period of time for project use. TA providers may assist stakeholders in other service provider models to develop technology systems for capturing and managing farmer profile data. Market research, monitoring, and impact assessments are common.	For-profit companies, remote data collection, business intelligence (BI), and MNOs collect farmer profile/ user-profile data to direct services or to sell data to other providers.
Examples of service providers that fall into the model			
International Center for Tropical Agriculture (CIAT), CGAP, Indian Agriculture Research Institute (IARI), Corpoica (Colombia), RAF Learning, Insight2Impact (i2i), MIX Market, GSMA, The Technical Centre for Agricultural and Rural Cooperation (CTA)	Ministries of Agriculture, delos Andes Cooperativa (Grameen Foundation partner), Philippine Coconut Authority (Grameen Foundation partner), ISAP India, Ghana Commercial Agricultural Project (GCAP), Olam/World Cocoa Association, FUNDER, Colombia Coffee Federation, One Acre Fund, Food and Agriculture Organization (FAO) Digital Services	Grameen Foundation, West Africa Agricultural Productivity Program (WAAPP), ADVANCE II (ACDI/VOCA), Mercy Corps' AgriFin Accelerate, Nathan Associates, Data-Impact/ISEAL, Mars	Example of B2B: Gro Intelligence, aWhere, ESOKO Isyt, GSMA, MNOs, Accenture Connected Crop Solutions, iShamba, MIX Market/One Acre Fund, Famerline, Viamo, NAPL Examples of B2C: MNOs, iShamba, ESOKO Tulaa, Amtech Technologies Ltd., Musoni, Farmerline, FlyBird Innovation, Sanjhi Tokri, Precision Agriculture for Development, FUNDER, Ricult, Akorion, One Acre Fund
Case studies that represent the model (see Section XIII of this document)			
CIAT/CGIAR	Grameen Foundation		Ricult

To ensure that data brings value to farmers and that the digital technology being promoted will ultimately be demanded by them, farmers should be integrated into the design process from the beginning and their perspectives taken into consideration.

The role of the farmer in directly managing his/her data is growing.

Given smallholder farmers' uneven ownership of digital technology, limitations in internet connectivity, and varying rates of literacy, farmers' direct use of their own data has been limited. Traditionally, farmers engaged with their data, if at all, using paper-based "passbooks" that recorded agriculture data from their transactions with an extension agent (Kilimo Trust, n.d.) or through a calendar to assist in tracking recommendations related to the harvest cycle (Mittal, Babcock, & Kienzle, Forthcoming). Service providers have requested data from farmers to understand their needs and capabilities; to design farmer products and services; and to monitor and evaluate changes over time. For example, agriculture extension agents might collect farmer profile data and use it to provide specialized support and recommendations to the farmer.

Francis Arthur, Grameen Foundation

More recent innovations, however, are leading farmers to interact more with their own data.

Digital Green, for example, is known for working directly with farmers to develop local videos on various agricultural or nutritional practices. Video is one way farmers can directly engage with "data" being collected on the practices of peer farmers.

The Feed the Future Ghana Agricultural Development and Value Chain Enhancement (ADVANCE II) project implemented by ACDI/VOCA, for example, used innovative, locally-sourced farmer identification (ID) smartcards (Ramasinjatovo, 2017). Each smartcard, owned by the farmer, contained the farmer's photograph, ID number, and a chip that stores his or her private data and information (such as information on trainings attended and farm data). While the farmer could not directly use the data, the smartcard became a valued identification tool—especially for female farmers—since so many lacked a national ID. The smartcard facilitated ACDI/VOCA's ability to track farmer data and evaluate the project's impact ("USAID Ghana ADVANCE," 2017).

Given the growth of smartphone ownership among rural populations in some areas, farmers are increasingly using their own data, primarily through application-based data synthesis. For example, Akorion, an agricultural technology company based in Uganda, has a suite of tools that farmers can access, either through their own smartphones or in coordination with a digitally-equipped field agent. The application collects data on farmers, such as their bio-data, the GPS of their plots, and the crops they grow. It uses this data to create customized profiles that help farmers make decisions about the quantity of inputs they need and at what price and where they can sell their yields to make a profit.

Mahindi Master uses game simulations to enable farmers to learn about and experiment with recommendations about fertilizer inputs for their particular soil type. A farmers' basic data and crop modeling allows the farmer to make decisions after exploring the impacts of different input amounts, costs, and crop yield outcomes.

The FAO Digital Services team is also developing applications for direct use by farmers and extension agents. For example, in Rwanda and Senegal, the FAO has been working with local service providers to develop four applications they plan to replicate for new markets across the globe (FAO, 2018):

- Livestock App: called "Cure and Feed your Livestock," this application provides real-time information on animal disease control and feeding strategies
- Weather and Crop Calendar: combines weather forecast information, crop calendars, and alert systems
- AgriMarketplace: connects producers, traders, and consumers to facilitate trade and access to inputs
- e-Nutrifood: provides information on production, conservation, and consumption of nutritious foods

Thus, while historically farmers have been data providers and made limited use of their own data, new technologies are changing this. It is expected that farmers will soon be able to store their own information on a blockchain—a secure distributed immutable database shared by all parties in a distributed network where transaction data can be recorded—and give permission to access their data to whomever they please ("Blockchain for Development," 2017). This could provide farmers with the opportunity to control who uses their data and how (Glover, 2017).

IV. TYPES OF DATA COLLECTED

Summary: Farmer profile data

- There is agreement that comprehensive farmer profiles are needed to give precise support to farmers.
- New ways of collecting and aggregating data and applying analytics can reduce the amount of direct input needed from the farmer, thereby reducing data-collection fatigue.
- The agricultural sector lacks good data on women.
- Marrying plant science data with farmer data is a new frontier for improving productivity.
- Taking an ecosystem approach to data is necessary since it enables addressing the comprehensive needs of a farmer.

Key considerations

- When initiating capture of farmer data, service providers must assess what data they are prepared and equipped to collect. Information should have a clear connection to their organizational goals and the purpose of the collection activity (e.g., increasing farmer productivity, extending financial services, connecting to purchasers or input providers).
- Service providers should make a concerted effort to understand what data they are already collecting, what is being collected by others and how these can be leveraged and augmented. Seek partnerships where possible.
- Leverage existing ontologies that facilitate cross-organizational data exchanges.
- Envision and articulate a sustainability model. There remains significant uncertainty and flexibility is required. What, if any, opportunities are there to monetize your data? Would others pay for access to your data or its learning products? Are there revenue-sharing models with other partners or stakeholders and what are the associated ethical considerations?

This section explains different types of data collected by the four main categories of service provider.

The more data you have on a farmer, the more individualized and timely an intervention can be.

In November 2017, CTA, a member of the Global Open Data for Agriculture & Nutrition (GODAN) initiative, commissioned a study to understand the role of farmer organizations and cooperatives in the agricultural data ecosystem. They compiled and categorized the types of data points collected by service providers, primarily focused on Sub-Saharan Africa (Boyera, Addison, & Msengezi, 2017). Similar efforts have been made by others, such as the Smallholder Farmer Market Segmentation Toolkit created by the Southern Africa Food Lab (Kelly, Manderson, & Khoza, 2016).

In Table 2 below, the first two columns represent the findings from CTA's assessment of the types of data that service providers have been collecting, providing the most comprehensive farmer profile framework to date. The CTA report outlined both basic and extensive profile data points. (Only the basic data points have been provided in the table to conserve space.) Additional items uncovered through research for this report have been added in italics in column two. These items were not captured in CTA's basic or extensive profile data points. (Also, the CTA report focused on crops and not livestock, so some information has been added here to recognize the gap. The present report also focuses largely on crops, but many data points captured in a farmer's profile for crops would also be applicable for livestock efforts.)

As the CTA authors highlight in their report, not all service providers would necessarily collect all these data points, but a comprehensive profile facilitates more individualized support to the farmer.

The third column includes new data points and/or technologies that may change or disrupt how some data points have been traditionally collected. For example, use of MNO Know Your Customer (KYC) data or other digital identity (e.g., biometric) data could be used to establish the identity of a farmer instead of requesting this information during a data capture session. Applications of sophisticated data analytics that integrate GPS, satellite data, and insurance models with other farmer data may be able to assist service providers in promptly supporting farmers in time of crop loss (Adam Wills, Natalia Pshenichnaya, Guillaume Chorivi, Gilles Morain, & Jamal Khada, 2015).

The new data technologies outlined in column three are meant to be illustrative and not exhaustive. (In column three, methods of capturing some of the possible new data points are also included.) The CTA report also explores some of these technologies and other data sources that could be available for developing comprehensive farmer profiles.

Some of the concepts noted in Table 2 are explored in greater depth in other sections of this report.

Table 2. Digital Farmer data and data technologies

Data Categories	Data Points/Types	New Data Technologies or Data Points
Personal information	• First and last name • ID number • Male/female • *Poverty level* • *Food security level* • *Marital status (married, single, divorced, widowed, separated)* • *Family size* • *Decision-making role on the farm (plot owner, plot manager, labor/support) and in the household*	• Blockchain to immutably record identity as a uniform and portable unit • Satellite imagery to predict poverty levels • Mobile phone, application usage, and social media to establish identity and predict gender and person's occupation • Social media data • Call records to collect personal data
Communication	• Language preference • Phone number • Type of phone • Type of information farmer prefers to receive (on crops, markets, etc.) • *Mobile phone usage data*	• Outbound dialing and interactive voice response (IVR) to detect the language of the farmer (based on the menu chosen). MNOs predict language of phone user by their location and identify unique subscribers through analytics (even for those with multiple SIM cards) • Preferences noted by the farmer during profile setup in mobile phone applications drives personalized services immediately to the farmer • Call detail record analysis to track usage patterns
Location	• GPS coordinates of home • Physical address: village, street, number	• GPS coordinates are becoming integral for some services because they bypass need for physical address data • Longitude and latitude (minimum and maximum) coordinates to identify administrative location hierarchy
Farm details	• Land/farm registration numbers • Labor force/employees (# of people working/ paid to work on farm) • Equipment owned (planting, harvest, post-harvest equipment) • *Livestock (types and numbers)*	• Application Programming Interfaces (APIs) to connect farmer profile to national land registries (e.g., to verify land ownership)
Field information	• Location of plots • GPS of plots • Geo-mapping/geo-fencing • Size of field • Elevation • Soil conditions • Land title • Crop history (crops grown over time) • *Type of watering/irrigation sources*	• To drive individualized information, GPS of nearest village can be sufficient for much information; GPS of center of plot even better • Sensors to detect soil moisture; 'virtual' weather stations to increase precision of measures such as evapotranspiration • Mobile soil testing tools/satellite imagery to predict soil type • Satellite data to map plots

Data Categories	Data Points/Types	New Data Technologies or Data Points
Crop information	• Crops grown • Varieties grown • Seed types and amount used • Spacing of plants • Equipment used	• GPS location to predict crops grown in an area (based on local knowledge) • Collection of seed variety data to drive more accurate information to the farmer (for example, generating more accurate predictions of crop performance)
Production information	• Date of planting • Spacing of plants • Intercropping • Weather data (rainfall, temperature, hygrometry) • Yields (date of harvest, etc.) • Pest/disease attacks • Post-harvest (storage, sales) • Adherence to Good Agricultural Practices (GAPs): types of planting, fertilizer, pest control, harvesting techniques used	• Weather and pest data is real-time, early-warning data (and is generally integrated with farmer profile data through APIs) • Social media and other apps to capture a farmers' current challenges • Satellite data and sensors to estimate crop yields • Date of planting is critical for driving relevant real-time information to farmers • Apps that capture feedback given to farmers to track two-way communication
Financial instruments	• Account ownership (does farmer have account?) • Mobile account ownership • Remittances • Payments/cash transfers • Financial services providers who hold/facilitate farmers' transactions	• Apps or API to partner data to track actual transactions, amounts, related fees and costs
Credit	• Whether credit accessed, loan size, use of loan • Farm business plan details	• A range of data (social media usage, farm behaviors, phone usage) to generate alternative credit scores • Psychometric testing to assess willingness to pay • Real-time tracking of credit transactions
Insurance	• Fields (livestock) covered • Risks covered (and period) • Insurance company name • Cost • Amount repaid if one of the risks covered happens	• GPS, satellite data, and insurance models aggregated to support farmer in time of crop loss
Qualification/ Certifications	• Trainings attended • Certifications received • Monitoring of compliance to standards	• Smartcards to register farmers for trainings • Certification registrations to simultaneously provide visibility to multiple actors
Business information	• Cooperative memberships • Agribusiness linkages • Markets farmers are linked to • Sales prices	• Agribusinesses and cooperatives have access to farmer profile data through shared data platforms • GPS data to predict farmers' use of different markets • Input and crop sales transactions in real-time

New ways of collecting data, aggregating data, and applying analytics require less direct input from the farmer and lead to more real-time data flows.

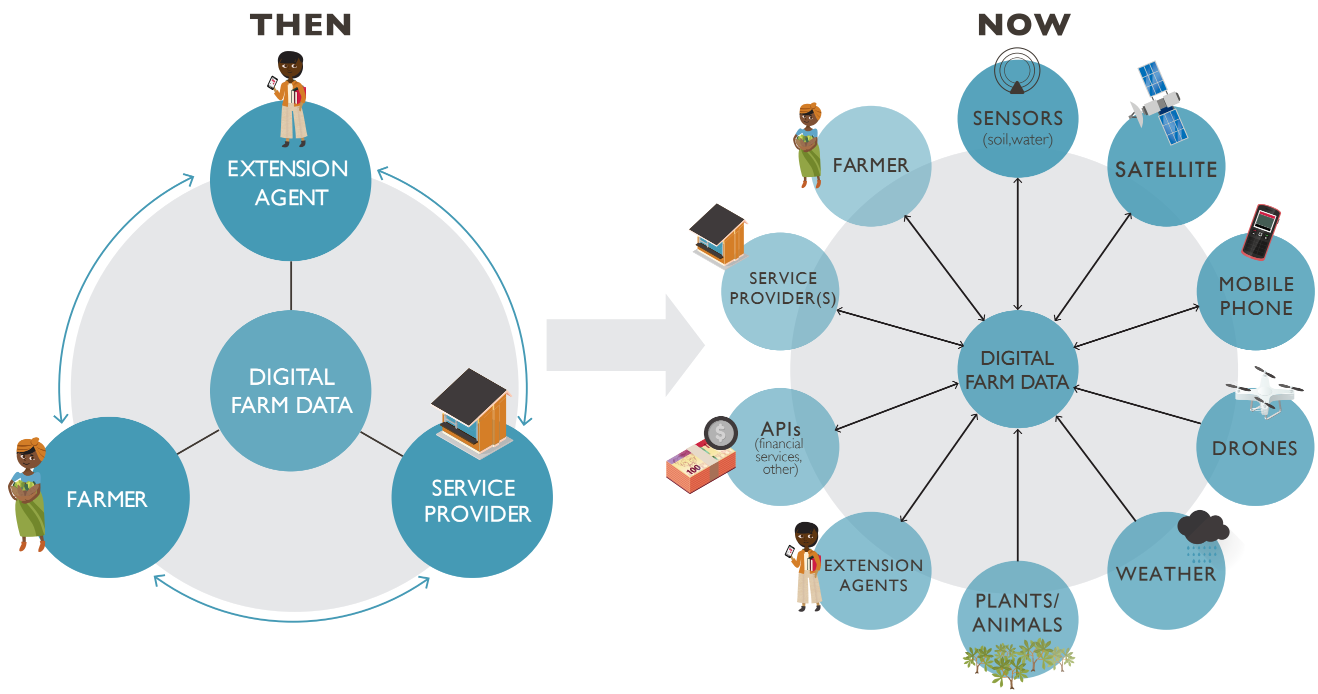
Jerry Glover of the USAID Office of Agriculture, Research, and Policy within the Bureau of Food Security noted that 'farmer fatigue' is a key challenge to collecting data from smallholder farmers (Glover, 2017). Historically, farmer/farm profiles typically relied heavily on data provided by the farmer, the extension agent, and the service provider. Farmers could be surveyed by multiple initiatives all asking essentially the same questions. Time is a precious asset on a farm; repeated surveys are a drain on farmers' resources.

Given advances in technology, today's data can come from multiple sources, easing the burden on the farmer (as illustrated in Table 2 above). Figure 1 below highlights the ongoing shift towards the farmer being but one key contributor to his or her profile and demonstrates the increased use of real-time data flows among the farmer, the service providers that support him/her, and the other technologies used by the farmer.

As digital management of farmer profile data becomes the norm, the farmer becomes only one of many sources of that data, and only one of its many users. As has been mentioned earlier, the farm becomes the common denominator for integrating data, not the farmer.

For example, with digital farmer or farm profiles, there are two-way data flows between the farmer and the service providers where data supplied by the farmer can support automated messages being sent by service providers back to the farmer. Also, when a farmer has sensors in his or her field to detect soil moisture, this data can automatically trigger an irrigation system or send a message to a farmer to alert him or her to the need to water the field.

Figure 1: Farm(er) data—then and now



For example, with the GSMA mAgri projects, the service provider (an MNO), typically collects only a discrete set of data, such as basic phone usage (Javed, 2016), the seed variety used by the farmer; whether crops are irrigated or rainfed, and the plot location (to determine whether it is in a dry or intermediate zone). This data is then combined with weather and satellite data to provide the farmer with individualized information.

Transporting Milk on a Bicycle in Rwanda, Ashley Peterson, Land O'Lakes International Development

FarmDrive in Kenya uses alternative credit scoring to improve financial inclusion. Its algorithm aggregates several sources of alternative data, such as market data, environmental data (crop portfolio, soil health, pests), social data (FarmDrive app usage), and individual data (demographic information, geographic information, plot size, production, and income). Some of this data comes from big external datasets, while the social and individual data comes from the FarmDrive app (Mensink & Vranken, 2017).

Despite the likelihood that less data is required from the farmer, several service providers have found that even while aiming to design a minimalist approach to collecting farmer data, development of the profile requires iteration, and this should be anticipated. For example, when the Mercy Corps AgriFin Accelerate team worked with the World Food Program to design Sibesonke, an application that aggregates smallholder farmers' produce to secure better sales prices for the farmers, it ended up needing more data than originally anticipated. The application initially relied on data points such as name, acreage/hectares owned, farmer group memberships, and the farmer's NGO partner. But the lack of data on gender and age limited their understanding of their farmer clients (Marita & Karlyn, 2018).

Gender information has been an untapped opportunity and new data analytics may help fill important gaps in knowledge and experience.

The use of ICT for agriculture (ICT4Ag) tools and data analytics has strong potential to help overcome gender inequities and to engage women more fully in agricultural development (World Bank, 2017). This potential remains largely untapped. Despite the existence of large agricultural datasets, little has been done in the development of data collection tools and analysis to understand gender dimensions (Hillesland, 2017).

A Special Note on Gender

One study conducted by Grameen Foundation in partnership with Farm Radio International on Grameen's ICT-enabled extension services (known as AgroTech in Ghana), noted the challenges to women's involvement in farmer groups (Asafu-Adjaye, Quaye, Yeboah, Osei, & Agbedanu, 2017). Increasing women's participation required intentional sensitization to overcome cultural restrictions. One approach worked with local village savings and loan groups to directly target women for agricultural extension support. This experience is consistent with many others: **integrating women and a gender perspective into agriculture takes intentionality** (Mbo'o-Tchouawou & Colverson, 2014).

Despite the proliferation of ways and means to collect data, data related to gender is still missing and this affects decision makers and policies (Urrea, 2017). Simply ensuring the inclusion of the sex of the farmer in a profile can provide interesting insights when coupled with sophisticated analysis (Quisumbing, 2014).

GENDER IN AGRICULTURE

Data2x's report also illustrates how satellite imagery can map rivers and roads, but it can also measure GENDER INEQUALITY *through integrating geospatial data with demographic and health survey data.*

Gender inequality remains a challenge for agriculture, in part because the sector lacks good data on women.

For a more sophisticated understanding of gender dynamics and social norms, both data collectors and analysts need to be gender aware and transformative in their approaches. This means considering indicators beyond just the farmer's gender. Indicators such as the farmer's decision-making power and control of and access to resources can provide crucial insights. It is not simply a matter of comparing male and female farmers, or male- and female-headed households. **We must understand how different household members behave, to analyze the difference in access to and control of productive resources, services, and employment opportunities** (Asfaw & Maggio, 2016). Feed the Future, IFPRI, and the Oxford Poverty and Human Development Initiative are leading some work in this direction, through development of the Women's Empowerment in Agriculture Index and its system of metrics ("Women's Empowerment," 2018). The FAO has also published guidelines regarding data collection on gender for agricultural projects. This publication highlights the importance of interviewing individual plot managers versus focusing on land owners or heads of household (Asfaw & Maggio, 2016).

Some interesting innovations in data capture methods are also being explored to analyze gender dynamics. For example, one project by CIAT will use accelerometers (GPS activity trackers) to generate big data on time use and energy expenditures—to serve as proxies for productivity and stress levels associated with different household and farm activities ("CGIAR Platform," 2017). Georgie Barrie from Farm.ink has written about their use of data analytics to understand gender. After data analytics revealed only 30 percent of their users were women, Farm.ink adjusted their marketing to better target women and increase women's participation in the system (Barrie, 2017). A recent paper by Data2X (Vaitla et al., 2017) also highlights how big data analytics can be used to understand women better—using social media to improve understanding of the mental health of girls and women and using cell phone and credit card records. The big data analytics highlighted in Data2x's report also illustrates how **satellite imagery can map rivers and roads, but it can also measure gender inequality through integrating geospatial data with demographic and health survey data to infer similarly high-resolution patterns of social and health phenomena across entire countries, such as differences in stunting between boys and girls or contraceptive use.**

Gender inequality remains a challenge for agriculture, in part because the sector lacks good data on women. Big data analytics could assist in filling this gap, but the process will have to be intentional.

Odetola Ismail Folaranmi, Feed the Future

Big data analytics may require less direct interaction with the farmer, but the precise questions asked become more critical. For example, to give very specific farming advice, information about the date of planting, fertilizer application, or harvesting is essential.

The greatest opportunity for increasing agricultural productivity globally is marrying plant science data with farming practices data.

This report has thus far focused on the kinds of data collected on farmers and how it is collected; actual information about crops harvested and seeds planted has not been discussed. In fact, one of the greatest opportunities for increasing agricultural productivity globally is marrying plant science with improvement in farming practices through precision agriculture (Thomasson, Santos, & Basu, 2016).

As one example, CGIAR has worked to develop more ontologies (or controlled vocabularies) for the digital annotation of plant science data, such as plant phenotypes and genotypes (The AIMS Team, 2016). They are also integrating some of this plant science data with farmer profile data (see short case study on Aclimate Colombia in section XIII).

Until now, however, **improvements in the agricultural sector have been incremental because many service providers focus on one factor—such as soil conditions—at a time, and not the ecosystem. Incremental change in just one factor cannot solve world hunger.**

Future progress can be made through the development of "more sophisticated analytical tools that can synthesize all forms of data" (Thomasson et al., 2016). This will enable the next step in optimizing farming practices.

While service providers may collect similar types of data, how they do so may differ.

The key informant interviews did not reveal significant differences in the types of data collected among the service provider models. While commercial actors tend to collect less data overall (which may be due to either their data capture methods or the business offerings driving their data collection), most appear to collect similar types of data. Their service providers may collect data directly or they may link to data provided by other service providers. However, the data capture methods used by the various service provider models present interesting differences. For example, commercial service providers tend to rely on mobile phones and applications, while governments and NGOs tend to rely more on collecting survey data through in-person interviews.

V. DATA CAPTURE METHODS

The methods used to collect farmer profile data run the gamut—from paper surveys conducted by field staff or researchers that are later digitized, all the way to data captures through remote sensing and through APIs that pull data entirely from other sources (such as national identification registers). The following section outlines how service providers capture data through traditional and new sources.

Digital data collection is the starting point for contributing to a digital ecosystem for farmer profile data.

While this landscape assessment focuses on digital data collection, several stakeholders interviewed were still using traditional pen-and-paper data collection and Excel spreadsheets to manage their data.

Connectivity, training, and mobile phone interface issues (among others) continue to challenge the uptake of mobile data collection tools despite their promise of reducing costs and improving data quality (Fitzgerald & FitzGibbon, 2014; Trucano, 2014). In the authors' experiences, service providers who rely on paper-and-pencil data collection do so for reasons related to cost, lack of knowledge and experience with tools, and the absence of a digital strategy (including a cloud-based management strategy, mentioned later). Firms often rely on staff using their own smartphones as one means to begin digital data collection. And ownership of smartphones is slowly growing.

Creating incentives or support for developing a digital strategy would be an important starting point for many, particularly government and civil society organizations. Many of the commercial service providers included here already use technology as their primary platform for collecting data and providing services.

Service providers are leveraging three main data capture methods: people facilitated, mobile-phone facilitated, and remote sensing/remote capture.

Some service provider models favor certain data capture models over others, but it is anticipated that over time these distinctions will lessen as the technology becomes more ubiquitous and if costs lower.

Summary –Data capture methods

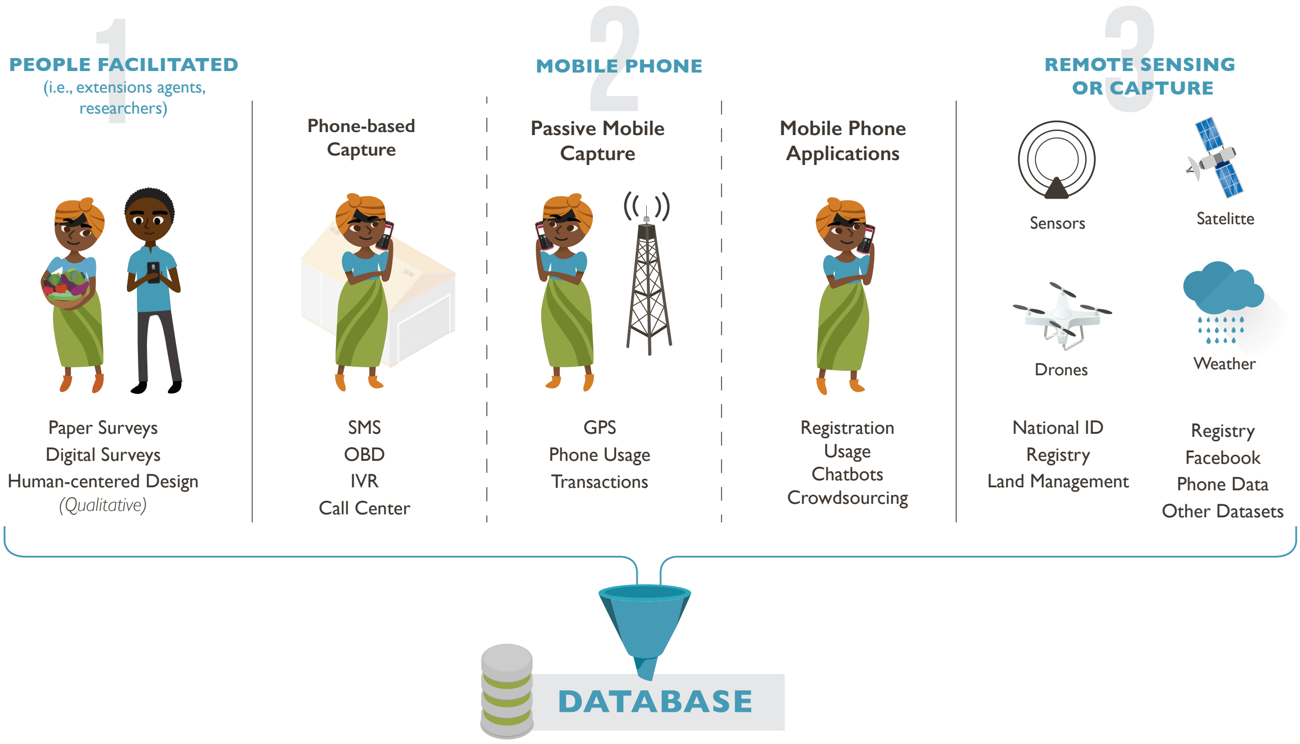
- Digital data capture is the starting point for developing a digital farmer profile.
- Service providers capture data with three main methods: people-facilitated, such as between an extension agent and farmer; mobile-phone facilitated, which captures data directly and indirectly from the farmer's own mobile phone; and remote-sensing or remote capture, which collects data from sensors, satellites, weather stations, or through connections made by APIs.
- More and more, farmer data can be captured without the farmer's direct involvement; for example, through use of sensors, satellite, drones, and API-facilitated access to other relevant data sources.
- Leveraging all three methods of data capture will enhance farmer profiles in the future.

Key Considerations

- What methods does your organization currently rely on for capturing farmer data? Do you collect data digitally as an important starting point?
- What technologies could you/should you consider when developing farmer profiles?
- Pay attention to the fundamentals of connectivity and usability. These key barriers are often overlooked or minimized. What is the mobile-phone ecosystem where you operate and how might mobile phones be used for data collection?
- New technologies such as satellites, drones, and sensors still face implementation challenges, but they can supplement or replace some data requirements and reduce farmer involvement in data capture.
- What other data sources (such as national information databases) might assist your organization in developing full farmer profiles? What B2B actors might your organization partner with for data sharing?

Figure 2 below summarizes the various methods used to capture data from farmers. In-depth descriptions follow.

Figure 2: Data capture models



People-facilitated data capture: The most well-known and traditional collection of data is by extension agents, field workers, or researchers working directly with farmers and using quantitative surveys or human-centered design (HCD)/ qualitative research.

Technologies such as ODK-based applications (CommCare, TaroWorks, SurveyCTO, Dooblo's SurveyToGo, Esoko's INSYT, Amtech Technologies, Source Map, or Olam's Farmer Information System (OFIS) platform and other proprietary systems) are used to collect data digitally for these person-facilitated surveys.

An example of qualitative research and HCD is provided by PHB Associates, which collaborates intensely with farmers and other stakeholders to develop archetypal farmer profiles. Service providers use these archetypes to design products and improve user experiences for certain segments of the population and to improve uptake of products and services designed for farmers.

Among the service provider models, people-facilitated data capture is most favored by researchers, government or NGO extension programs, and technical assistance/project-based providers.

Mobile phone-facilitated data capture: The farmer's own phone is used to capture data. Three mechanisms can be used:

- **Phone-based data collection:** Farmers are contacted through SMS, IVR menus, outbound dialing (OBD), or chatbots that collect data directly (via an app) or indirectly (from Facebook and social media posts).
- **Passive data collection:** The mobile phone's GPS location or movement and usage data (such as amount of airtime used) are collected.
- **Applications designed for use on the farmer's phone:** These collect data directly from the farmer as he/she creates a personal profile for the app or indirectly as others participate in crowdsourcing. This method of data capture is most favored by commercial service providers, such as iShamba (based in Kenya) and Akorion (in Uganda). Akorion has developed an app called Ezyagric that can be downloaded by both farmers and extension workers. It assists farmers in mapping their plots, managing their data, buying and selling produce, buying inputs, and purchasing soil testing—all via the phone.

MNO Dialog Sri Lanka initially provided farmers with a phone number to dial into and navigate through an IVR menu to register for their agriculture information service (Palmer & Darabian, 2017a). However, too many incomplete registrations prompted a switch to OBD. The system could infer a farmer's language preference by identifying the language of the OBD to which they responded when asked to choose their preferred crop. This resulted in a **"one-touch" registration.** After this registration, more OBD requests were sent to gradually complete the farmer's profile. Questions focused on seed types, agro-climate zones, and irrigation methods to further customize the farmer's content offerings. **This approach allowed farmers to schedule the data collection calls at times most convenient to them, resulting in greater farmer participation and satisfaction.**

In addition, **this sort of data capture can be used to create default profiles with just a few pieces of data, allowing farmers opportunities to correct or adjust them.** For example, a farmer may be asked to provide his or her land size and preferred crop. Later, an OBD call can verify the data with the farmer. After receiving information via SMS or IVR, the farmer can be invited to push a number to change his or her location, crop, and other details as needed.

Applications can also collect data through crowdsourcing. An application being tested in China called eFarm tries to marry remote-sensing data collected via sensors and satellite with "human-sensing" data (Yu et al., 2017). The eFarm app pulls down volunteered geographic information (VGI), such as one might find on a website like Geo-Wiki, and crowdsourcing data (such as data one might find from an online farmer self-help group) to contribute to timely and geo-referenced farm management information that can be linked to spatial land parcels from remotely-sensed images. Crowdsourced data, however, has limitations. While research has shown that both experts and non-experts can identify human impact on land, experts are more likely to identify land cover type accurately (See et al., 2013).

Pokhriyal and Jacques highlighted in their study the promising use of OpenStreetMap data, which garners crowdsourced pictures of facilities and roads that can be used for more near real-time imagery of locations. Whereas one would need several years of satellite data to detect change over time, for example for changes in vegetation, crowdsourced pictures could provide immediate imagery of changes in vegetation. The researchers also tested whether mobile phone and satellite data could be used to predict poverty levels in Senegal. They found that relying on mobile phone-based data alone can leave out important demographic groups such as youth, the ultra-poor, and women, due to poor electrical and network connections (Pokhriyal & Jacques, 2017).

Companies such as Digital Globe promote crowdsourcing as a means for data collection, particularly in parts of the world that do not have reliable or readily-available maps ("DigitalGlobe," n.d.)

Remote sensing/remote capture: This method includes data collection via sensors (such as for soil moisture or for tracking cattle), satellite, weather stations, drones, or APIs.

Sensors

The Internet of Things (IoT) is a term that has been coined to describe how sensor data is collected, transferred (generally through the Internet or cloud) and then made available in a database for analysis and use. Almost any physical item—medical device, plant, watch, light bulb, security system, and so forth—can be transformed into an IoT device if it can be connected to and controlled via the Internet (Ranger, 2018).

Business Intelligence (Newman, 2017) estimates that there will be 55 billion IoT devices by 2025, up from about 9 billion in 2017, and nearly $15 trillion in aggregate IoT investment between 2017 and 2025. Use of IoT is accelerating, but challenges to harnessing this technology remain. These relate to power sources and connectivity; the calibration and cost of sensors; training on their use; network bandwidth as more devices are connected to internet (Biggs, Garrity, LaSalle, Polomska, & Pepper, 2016); and cyber security threats (Manning, 2016). In this document, as well as in the literature generally, examples of sensor use are limited to collection of data about the weather and soil moisture (Karim, Anpalagan, Nasser, & Almhana, 2013). In developing countries, documented challenges regarding the use of sensor data point to the cost and reliability of sensors and the capability of farmers to use the sensor systems, among others (Karim et al., 2013; Lee & Choudhary, 2017).

FLYBIRD INNOVATIONS: A MINI-CASE

Flybird Innovations is a social impact agricultural enterprise in India supported by Chennai-based Villgro, a social impact fund backed by Dell Foundation. It has developed a product known as Siri, which is a smart irrigation controller. Siri manages both water and fertilizer application from sensors that gather data on soil moisture, temperature, and humidity—with the goal of preventing both under- and over-irrigation and fertilization of crops and plants. The product can be programmed based on a timer, based on volume of water and fertilizer released, or based on information gathered by sensors (with sensor-based being the most expensive). It is programmed to resume monitoring following any power outage. When used in greenhouses, it can also control heaters, coolers/fans, and irrigation ("Flybird Innovations," 2014).

Besides sensor data, Flybird collects basic demographic information on its farmers and geographic and crop data. This allows Flybird to predict water requirements and optimal fertilization for farmers' crops (KS, 2017; World Bank, 2017). A farmer can monitor his or her data through either a mobile-based app or a computer-based app. Flybird claims Siri enables farmers to achieve a 15 percent to 20 percent increase in their yields and a return on their investment within six months (due to reductions in power, labor, and water costs).

Flybird Innovations faces these challenges with smallholders (see box). A sensor-based irrigation system can cost upwards of $400. While smallholders are interested in the technology, they are requesting subsidies, which is a possibility Flybird is exploring ("Flybird Innovations," 2014).

Drones

Drones, sensors, and IoT often go together. Agrilift is a company that is piloting crop monitoring technology using drones. Their technology takes overhead images of crops at different time intervals. It analyzes these images with open-source models of plant growth that can identify nutrient deficiencies and diseases and provides information on soil conditions. This data can help farmers determine which seeds to plant and the amount of fertilizer to apply (Mensink & Vranken, 2017).

At this stage of its development, several factors make the use of drone technology for farm data collection challenging (Afadhali, 2016). These include: 1) the weather, since drones cannot fly during the rain or high winds; 2) the size of farm plots, since the smaller the plot, the harder it is to extract accurate data; and 3) the difficulty of interpreting drone data (making smallholders dependent on field agents or other experts to share information).

Satellite and weather data

Unlike sensor and drone data, satellite data is often collected by government agencies for public use (Shekhar, Schnable, LeBauer, Baylis, & VanderWaal, 2017). For example, the National Oceanic and Atmospheric Administration and the National Aeronautics and Space Administration (NASA) make available huge volumes of satellite imagery with increasingly high resolution and at great frequency. Many open imagery datasets are accessible through platforms such as Amazon Web Services S3, Google Earth Engine, and NASA Earth Exchange. Companies such as Digital Globe, Earth Networks, and aWhere help make satellite and weather data useful for practitioners since this data is often difficult to interpret and integrate into day-to-day use for businesses, ensuring greater use of this data for the benefit of farmers.

Those who have been piloting the use of satellite data have noted several challenges including imagery resolution, cost, and availability of expert knowledge. For example, Geodata-Based Information Services uses spatial and other geo-data to provide time- and location-specific advisory services to smallholder farmers in Bangladesh. They have found that small plot sizes pose a challenge for satellite imagery resolution (Mensink & Vranken, 2017). This is consistent with findings from available research (Burke & Lobell, 2017). Hiran Bhadra of the Accenture's Digital Agriculture Service team was interviewed for this report. He shared that the cost of satellite data can be challenging:

> *"We take the satellite data, it is an important data source for the state of the crop. You can get to 15 centimeters (cm) of depth, you can do 30 cm, but the cost difference between 15 cm and 30 cm can be in the hundreds of thousands of dollars in cost. A 15 cm view may be critical, but it comes at a cost. The costs will become more efficient over time, but to employ the vision of using this new technology, we need to be intelligent about [satellite's] use."*

APIs

This category of data collection covers any data that can be collected to complete or enhance a farmer's profile through an Application Programming Interface (API). In many cases, an API is necessary to integrate weather forecast and satellite data, but an API may also be useful to pull data from other data sources, such as a national identification register. For example, aWhere is a commonly-used B2B company for capturing satellite data, utilizing soil moisture models, conducting weather forecasting, and gathering pest and disease risk information. This data is often connected to a business' own data through an API ("Agricultural Weather," 2018).

In the future, farmer profiles would benefit from all three methods of data capture.

Any one of the three main methods or combinations of them can be used to feed data into ever larger datasets and benefit from more powerful levels of data analytics. In the future, robust farmer profiles can be developed by utilizing all three methods, reducing input required by the farmer. Ricult, a start-up based in Pakistan, draws on multiple sources of farmer profile data. (See further information about Ricult in Section XIII on case studies.) They collect their own data via field agents equipped with android tablets and also use proprietary psychometric testing to determine willingness to pay for approving access to their credit services. This data is combined with SMS data from others who have transacted with a particular farmer, satellite data on soil and crop analytics, weather data, and information from other databases that help establish identity. Ricult CEO Usman Javaid was interviewed for this report. He shared, "Most of the data we need on a farmer can be collected remotely."

The choice of data capture tools is often driven by the cost and the availability of the technology. For example, having field agents equipped with smartphones or tablets can determine the feasibility of any digital data collection. Some companies' services also include providing technologies for capturing farmers' digital data. These include Esoko INSYT, iShamba, and Akorion. Service providers who are not currently leveraging digital technology note the difficulty in identifying appropriate technologies and covering the costs of these investments.

Photo by Rhiannon O'Sullivan, World Vegetable Center

VI. DATA STORAGE: CLOUD COMPUTING AND BLOCKCHAIN

This section outlines how cloud-based services and blockchain are increasingly underpinning data storage and processing. Cloud-storage is currently favored more by commercial actors and those service providers who have incentives to share data—such as technical assistance providers and their partners.

Cloud-based services are a significant piece of architecture that makes big data analytics possible.

Cloud-based services to store, process, and analyze data have been available for some time and they remain among the most significant pieces of architecture needed for big data analytics (Hashem et al., 2015). Cloud storage centralizes data remotely and facilitates the use of files and other software by multiple users without their having to store data on their individual hard drives. While cloud-based storage and computing is not essential—many organizations rely on their own servers for storage space—cloud storage and computing help blockchain, digital wallets, lending platforms, and the IoT work. Cloud storage facilitates integration and analysis of newly-captured data with that of data already available in the cloud. The top companies providing cloud storage and computing services are Amazon Web Services, Microsoft, Google, Alibaba, IBM, Oracle, and SalesForce (Rosenberg, 2017). New companies entering this space are CenturyLink, Virtustream, Rackspace, and Fujitsu.

But cloud-based data management comes with a cost, and a few of the organizations interviewed—particularly smaller non-profit organizations—noted that despite their desire to use the cloud, the cost is prohibitive. While the cloud is promoted to reduce infrastructure and maintenance costs associated with big data analytics, it is not affordable for many small and medium enterprises, particularly given that "big data analysis requires a huge amount of computing resources." It also introduces new security risks ("Big data changing," 2014).

Summary –Cloud-based storage and blockchain

- While cloud-based storage and cloud-based computing are not essential, these help blockchain, digital wallets, lending platforms, and IoT work. They facilitate the integration and analysis of newly-captured data with data from sources already in the cloud.
- Today, blockchain relies on cloud storage and cloud computing; in the future it may underpin cloud storage.
- Blockchain promises to facilitate digital identity, digital value transfer, and monetization of farmer data by the farmer. It also holds promise for reducing corruption.

Key Considerations

- Does your organization currently use cloud storage? Cloud computing? Why not?
- How might your organization benefit from utilizing cloud storage? What data might you access with cloud storage and computing?
- The models for data processing and storage have implications for risk, privacy, and costs. Different models may be needed for pilot and scale.
- What benefits might blockchain technology offer your organization?

The team at TaroWorks (Chau, 2017) noted there are three main cloud types: *public clouds* like data stored on Salesforce (upon which TaroWorks is built); *private clouds* like those facilitated by Microsoft (data is stored on an organization's own servers but is made available within an internal network via the internet), and *hybrid clouds* that utilize both public and private servers. TaroWorks also noted four decision points for cloud computing services:

- cost (for software and hardware as well as expertise needed to manage the services)
- connectivity (network bandwidth issues)
- collaboration (who needs access to the data and at what levels of security)
- data complexity (how much will be stored and how complex the analytics are)

Data complexity is a particularly important factor when considering cloud storage for cloud computing with advanced analytics. Service providers must determine how big their data requirements are so that enough cloud space is rented to ensure big data analysis can be conducted to the fullest extent (Faleide, 2017).

Blockchain relies on both cloud storage and cloud computing and may soon underpin cloud storage.

Many if not all people contacted for this assessment had heard of blockchain but very few had actual experience with it. And while conversations did not focus on cryptocurrencies, these two technologies are often mentioned in the same breath (Faleide, 2017). Blockchain is a "secure distributed immutable database shared by all parties in a distributed network where transaction data can be recorded" (Corea, 2017). Transaction data is permanently recorded in files called "blocks." In simpler terms, blockchain can be imagined as Google sheets (versus just Excel spreadsheets) shared through email. All those with access to the Google sheet and with adequate permissions can have real-time access to the data and can track every update being made. Once entries are made, they are recorded permanently and cannot be edited or erased by anyone (Sanghera, 2018). Blockchain may eventually underpin cloud storage (Mearian, 2018). Recent articles point to the development of blockchain-based distribution storage built on large amounts of unused storage on hard drives of people all over the world (and helping members of the blockchain monetize—or sell—their available storage space to others on the blockchain) (Puranik, 2018).

Blockchain facilitates data sharing and creates potential for farmers to own and manage their own data.

Blockchain lets people and organizations share information (e.g., land rights, identification, ownership of assets, transactions) securely and transparently and in a decentralized manner that mitigates the need for any centralized control by private or government entities. This technology may eventually help development assistance providers avoid corruption in developing countries. According to the GSMA, blockchain can be used to help people protect vital information and keep it from falling under the control of any one entity. Information may include land rights, proof of citizenship, or the official identity records of refugees crossing borders ("Blockchain for Development," 2017). According to Corea, "It is undeniable that artificial intelligence[2] and blockchain are two of the major technologies that are catalyzing the pace of innovation and introducing radical shifts in every industry" (Corea, 2017).

Blockchain could also enable farmers to securely store their own information and give permission to access their data to whomever they please ("Blockchain for Development," 2017). According to Jerry Glover, USAID Office of Agriculture, Research, and Policy (who was interviewed for this report), this could provide farmers with the opportunity to monetize access to their own data as well as avoid the fatigue associated with repeated data collection visits. At least 1.1 billion people worldwide currently lack any form of official identification—particularly rural residents, women, and children ("Blockchain for Development," 2017). Blockchain could be the platform for "user-centric" ID systems that store data such as first and last name, phone number, date of birth, gender, and nationality. Glover says farmers could also use blockchain solutions to store and authenticate personal data such as their income, transaction histories, credit worthiness, geo-location, farm size, etc.

2 Artificial intelligence is the end result of the increasingly sophisticated stages of data analytics: diagnostic, descriptive, predictive and prescriptive.

BanQu is a company that uses blockchain to provide people with portable digital identities. Anyone with a mobile phone can set up their economic identity on BanQu. The company is currently piloting a project in Latin America that uses small-plot farmer land mapping to facilitate farmers' access to finance (especially women farmers). Ashish Gadnis of BanQu was interviewed for this report and shared how blockchain brings a level of reliability to agricultural value chains where the farmer and the other members of the blockchain can trace and produce evidence for every bag of produce for every agricultural cycle. Farmers benefit from this traceability as it builds historical evidence of their product, the amount they were paid and by whom, etc. If the buyer were to ever go out of business, the farmer would still have their own records, which could be shared with other service providers for acquisition of services. Ashish believes that "data collection will change in the next 18 to 24 months" as more granular levels of data are added to blockchain and become available for predictive analytics.

Blockchain, however, can only function where there is a consistent internet connection. Before blockchain can be adopted by smallholder farmers and service providers that support them, the rural infrastructure required for digital transactions must be in place ("Blockchain: Beyond Bitcoin," 2017). Gro Intelligence posits that blockchain could transform agriculture, but this lies somewhere in the future, and there will be many pitfalls. Usman Javaid, CEO of Ricult in Pakistan (also interviewed for this report), believes there is a bright future for blockchain and digital currencies:

"I see a future where you'll have a lot of digital currency driving rural economies. I believe that there will be a cashless economy that will be fueled by blockchain and digital currency. Uptakes of mobile wallets are low in Pakistan. If a farmer needs to get cash somewhere, this isn't convenient. [Blockchain will result in] an empowered farmer. He drives the data that is provided. He monetizes it, uses it for his benefit."

Photo by Thom Wallace, Africa Lead

VII. DATA ANALYTICS

This section outlines data analysis tools and defines the concept of "analytics" and the types of analytics service providers can use to make use of smallholder farmer profile data. Commercial service providers and researchers who leverage big data sources tend to favor more sophisticated analytics. Some commercial service providers have also based their business models on the use of analytics to provide services to farmers.

Data analysis tools

Those interviewed for this assessment mentioned numerous data analysis tools. These included TaroWorks, SharePoint, Oracle databases, MERGDATA, DATAWINNERS, SMARTEX, and Amtech Technologies' EasyMa data management system. Some companies, like FlyBird Innovations, have developed their own remote-sensing data capture and analysis technologies to transfer information to decision-makers. Any one of the service provider models can make use of these tools to analyze data.

Defining analytics

In 2016, Krista Nordin of Insight2Impact highlighted the elements of a data value chain for financial services providers (Nordin, 2016) and these elements can serve as a model for the agriculture sector. As Nordin notes in her paper, **"companies are winning or losing depending on how they are using their data."** In highlighting the potential of analytics, the following section draws heavily on Nordin's paper as well as others regarding types of analysis being used to create value from farmer data.

Summary: Data analytics

- Data analytics are being used to create "new data" from existing data.
- Service providers are "winning or losing depending on how they are using their data." Increasing use of sophisticated analytics (such as predictive, prescriptive, and cognitive analytics) is changing how data is captured, used, and shared in real-time.
- No two environments are exactly alike and the development of strong analytics is more laborious in agriculture than other sectors.

Key Considerations

- How does your organization currently analyze its data? Does it use descriptive, diagnostic, predictive, prescriptive, or cognitive analytics?
- How might any of these types of analytics improve your organization's current approach to data analysis? What seems realistic? What is imperative?
- What data science/data analytics skills currently exist within your organization? What gaps in skills exists? What skills should be acquired internally by staff and what outsourced to consultants?
- There is a progression associated with the application of analytics. While it is not necessary to utilize all analytics methods, improved efficiency and applicability can increase relevance for and buy-in of smallholder farmers.

Five categories of analytical methodologies are discussed here: descriptive, diagnostic, predictive, prescriptive, and cognitive. These categories represent increasing levels of complexity, data volume, and data type requirements. Nordin's research estimated that 99 percent of businesses use descriptive analytics, but that by 2025, businesses will spend more than $60 billion on cognitive solutions (Nordin, 2016).

Descriptive analytics look to the past to discover "What happened? When? Where?"

Descriptive data are often included in real-time dashboards and analyzed with basic statistical methods. Most service providers have had experience using descriptive analytics to monitor progress over time or to conduct basic program research or evaluation.

CGAP'S SMALLHOLDER FARMER DIAGNOSTIC ANALYTICS: A MINI-CASE

The first phase of segmenting smallholder farmers in CGAP's smallholder farmer diary research involved a machine learning algorithm called Random Forest. It assessed which indicators correlated most closely with high rates of financial account ownership among farmers. The six measures were educational attainment of the head of household, poverty status of the head of household, access to emergency funds, mobile phone ownership, attitude toward the future, and encountering unexpected life and farming events. Using these variables, CGAP identified five clusters of farmers:

FARMING FOR SUSTENANCE: *Farming households that struggle to support their needs and have the most to gain from financial and agricultural products and services.*

BATTLING THE ELEMENTS: *Farming households that have persevered through significant agricultural challenges and remain optimistic toward farming. Financial mechanisms have enabled some of their perseverance.*

DIVERSIFIED AND PRAGMATIC: *Farming households that grow more, sell more, earn more, and have more income streams and connectivity to financial mechanisms. Even though they take pride in farming, they would consider diversifying out of agriculture if given the choice.*

OPTIONS FOR GROWTH: *The youngest group of the five, these farming households rely heavily on agricultural income but are also the most likely to have more stable income sources outside of agriculture. They could stay or leave farming, depending on opportunities.*

STRATEGIC AGRICULTURAL ENTREPRENEURSHIP: *Actively engaged in agriculture and aspiring to grow their agricultural activities, these farmers could inspire other farmers. They have access to more resources, savings, etc., that support their resilience to agricultural shocks.*

Source: Anderson, J., Marita, C., Musiime, D. (2016).

As described in the box below, CGAP uses *descriptive statistics* to summarize the details from smallholder farmers' diary data but also uses *diagnostic analytics* when applying the Random Forest algorithm to classify farmers (Anderson et al., 2016).

Diagnostic analytics look to the past to discover "Why did it happen?"

Descriptive data can be subjected to statistical regression and the results can be viewed on analytic dashboards.

Grameen Foundation uses TaroWorks (whose analytics engine is powered by Salesforce) to develop customized dashboards for different types of service providers. The dashboards provide both descriptive and *diagnostic analytics*. For example, they provide dashboards on who is improving their yields by adopting GAP—with data by region, by gender, etc. The dashboards also classify whether and which farmers have reached certification standards based on a series of criteria (such as GAP adoption). The analytics are portrayed in simple charts, graphs, and gauges that help partner staff identify the high performing farmers as well as those who require additional assistance (see Figure 3).

Photo by Bimala Rai Colavito, ENBAITA

Figure 3: Grameen's use of TaroWorks for crop certification and GAP adoption monitoring.

Source: Grameen Foundation

Predictive analytics look to the future to determine "What is likely to happen?"

Predictive analytics require larger volumes of data and computing power. Below describes three different use-cases of predictive analytics: alternative credit scoring, yield gap analyses, and use of integrated data sources to support farmers with actionable data.

One of the first sectors to embrace the use of *predictive analytics* was the insurance industry. According to the Society of Actuaries, the sector has been "predicting using data for a long time, but now [it] can do it utilizing exciting new tools, new sources of information, new techniques, and...at a more granular level than ever before" (Trachtman, 2017). It involves "scrubbing and analyzing the raw data and augmenting these data with relevant external data," such as satellite and sensor data, among others (Andrews, 2017). This augmentation can be referred to as the use of "big data." Therefore, **the concepts of *big data* and *predictive analytics* are often used interchangeably.**

Many proprietary and freeware off-the-shelf software packages assist in developing predictive analytics. These include SAS Predictive Analytics, IBM Predictive Analytics, and RapidMiner, among others ("What is predictive analytics?" 2018). Many of the service providers interviewed for this assessment who used predictive analytics have developed their own proprietary analytics. Some, like One Acre Fund, have leveraged R programming language for in-house work or outsourced the analytics to private companies. One Acre Fund collaborates with the analytics firm Quantitative Engineering Design.

Alternative credit scoring

The financial services sector has also leveraged predictive analytics for credit scoring. Financial institutions (FIs) rely on past client behaviors to predict future behaviors of both new and existing clients. With the emergence of big data and access to additional data sources, FIs can incorporate alternative credit scoring models, particularly in countries where credit bureaus are not active or when people are unable to demonstrate past financial behaviors (such as prior use of credit). In Kenya, the Commercial Bank of Africa worked with MNO Safaricom to develop a nano-loan product. Given the lack of access to information on credit behaviors, they developed a scorecard that drew on mobile phone usage, such as top-up frequency, as a predictor for capacity to repay (Caire et al., 2017).

To increase access to agricultural financing, **service providers are also drawing heavily on predictive analytics to determine farmers' willingness and capacity to repay loans.**

For example, Grameen Foundation has been testing a similar alternative credit scoring methodology called the Agricultural Risk Evaluation Tool (ARET) (Tobias, 2016). With this tool, Grameen aggregates and merges data sources available to its partner in Colombia, uses chi-square, correlation, and t-tests to identify those variables significantly associated with loan repayment, and uses the Classification and Regression Trees (CRT) algorithm to build a risk model. The current prototype includes seven variables (such as on-farm behaviors, farm equipment ownership, and income) to classify farmers' risk. This project is still underway and variables in the algorithm are still being adjusted. Many financial and nonfinancial services providers who develop alternative credit scoring or risk methodologies finetune their models on an ongoing basis (Caire et al., 2017).

Ricult in Pakistan uses predictive psychometric tools to assess the personality characteristics and cognitive abilities of their applicants to gauge willingness to pay. They use credit scoring algorithms that rely on both traditional and alternative sources of data to determine ability to repay a loan.

Lenddo and EFL Global are two B2B companies who recently merged to provide alternative credit scoring services. They provide a product tour of their credit scoring process at https://producttour.eflglobal.com.

Yield Gap Analyses

A 2017 case study documented One Acre Fund's journey of integrating digital tools into their practices (Manfre et al., 2017). One of the team's aspirations was to integrate predictive modeling to better support farmers with personalized support. Towards this end, they have begun to undertake yield gap analyses in their work. In Kenya, they integrated their own farmer data (for approximately 4,500 farmer fields) with aWhere's remote-sensing derived weather data, NASA's Normalized Difference Vegetation Index, and Land Surface Temperature data to identify the most important agronomic levers to influence yields in different regions of operation ("Kenya," 2017). Using a Random Forest modeling approach, they discovered the most important yield predictors were location, weather, fertilizer application rates, field characteristics (e.g., size and distance from the homestead), biotic stresses and their management, and pre-season management of the field. The case study report documents outcomes related to agro-climatic zones that share environmental conditions conducive to different types of plant growth. In addition, One Acre is conducting prescriptive analytics to identify the types of recommendations for those living in different agro-climatic zones that should be designed for farmers.

Farmer Support with "Big Data" Analytics

The Aclimate Colombia project was a cross-sectoral partnership led by CIAT that engaged the Colombian Ministry of Agriculture and Rural Development, the National Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM), and Fedearroz (an association of rice growers). CIAT combined Fedearroz's data, which consisted of over 25 years of rice surveys, monitoring of harvest results, and plant science data with IDEAM's station-level weather data that was germane to determining rice growth. In 2014, the Aclimate platform aggregated and analyzed the data and predicted a major dry period would disrupt the upcoming growing season. The appropriate response to this prediction was to delay planting. Equipped with this prediction, Fedearroz informed rice farmers of the ideal time frames for planting. The analytics also revealed important regional variations in rice productivity. Some were mostly dependent on the average minimum temperature while others were mostly dependent on the amount of solar radiation or frequency of rainfall. These important and regionally-specific insights improved farmer decision making, resulting in higher yields compared to those farmers who ignored the Aclimate recommendations. A study conducted by the Open Data Institute to assess the impact of this open data platform estimated that improvements to data-driven decision-making by farmers averted potential economic losses of $3.6 million (Young & Verhulst, 2017).

Prescriptive analytics look to the future to determine "What should be done?"

Prescriptive analytics takes predictive analytics one step further and links the prediction with a prescription for what service providers and farmers should do or how they should respond to a particular situation.

One company, Ayata, promotes prescriptive analytics as "the combining of the science of predictions with the science of decision making" and notes five pillars that allow for prescriptive analytics (Figure 4):

Figure 4: Ayata's five pillars of prescriptive analytics

Source: Ayata Prescriptive Analytics, https://ayata.com/software/

The Mahindi Master project provides one example of the use of prescriptive analytics. It combines crop modeling, rainfall data, and soil information into a simulation application that "enables farmers to experiment and learn as they would on their own farm, without risking real investments in purchased inputs." Travis Lybbert, of the University of California, Davis, one of the masterminds behind Mahindi Master, shared that a soil-testing project in Kenya can produce one to two compact pages of data. However, this data is not digestible or actionable for smallholder farmers. Because one of the greatest known constraints to agriculture extension support is ensuring farmers see the value of information, just providing a recommendation might not lead to a farmer taking action—especially if it might incur a cost—and the information might not be trusted (Lybbert, 2018).

MARKET-LED USER-OWNED ICT4AG ENABLED INFORMATION SERVICE (MUIIS): A MINI-CASE

MUIIS is a mobile phone-based and bundled fee-for-service (agronomic, weather and insurance) offering that is delivered on Ensibuuko's MOBIS platform. MOBIS is cloud-based software that integrates with third-party solution providers by facilitating mobile money transactions and SMS alerts.

With the MUIIS service offering, farmers can receive automated alerts based on satellite data about the farm and on agronomic data provided by the consortium partners.

aWhere integrates the data from eLEAF (agronomic data enabled by satellite images of the farmer's field), EARS, Mercy Crops, and AGRA. aWhere integrates some of its own weather data. Using data collected by MUIIS service agents, Ensibuuko pulls the precise weather and agronomic data through aWhere's API.

The data made available through aWhere's API is ground-truthed by agronomists at AGRA and the National Agricultural Research Organisation (NARO). Once the data is interpreted and validated by the team of agronomists, messages are crafted with support from extension specialists based on the profiles of farmers and on the format to be prescribed by Ensibuuko.

Source: Building the ICT Infrastructure of MUIIS for SMS Delivery. (2017). MUIIS. CTA. http://muiis.cta.int/2016/12/21/building-the-ict-infrastructure-of-muiis-for-sms-delivery/

By drawing on actual soil data collected from the farmer and then combining this with weather data (rainfall, etc.) and farm spatial data, Mahindi Master computes recommendations for fertilizer use through a series of learning opportunities. Farmers receive the recommendations via an app and can test different fertilizer combinations to quickly learn how they will interact with their plots' characteristics.

Satellite and weather data are also often used both to predict and recommend action. aWhere is a satellite and weather company often used by other service providers to develop their own prescriptive analytic models, particularly Early Warning Systems for pests and weather events (Camp, 2016). One such initiative in Uganda that uses satellite services and prescriptive analytics is the Market-led User-owned ICT4Ag Enabled Information Service (MUIIS) (see box on the left). MUIIS uses satellite-based information services for weather forecasting and alerts, crop management, and agronomic tips and financial services (including index-based insurance for farmers) ("MUIIS," 2017). MUIIS is operated by a consortium of CTA, aWhere, EARS Earth Environment Monitoring, Mercy Corps, AGRA, the Eastern Africa Farmers Federation (EAFF), and the eLEAF Competence Center. MUIIS is owned by EAFF, with other partners having revenue-sharing agreements. This ensures that content continues to be relevant to farmer users.

Cognitive analytics look to the future to answer the question "What is the best action?"

Distinguishing between cognitive analytics and predictive or prescriptive analytics may be a false dichotomy. Some sources indicate that cognitive analytics adds the dimension of unstructured data (like data pulled from Facebook or social media) to structured data (Krishna, 2017). However, Ayata's pillars of prescriptive analytics (in Figure 4 above) note the incorporation of unstructured data. Several key informants said they use cognitive analytics while others felt that cognitive analytics is just an advanced use of predictive or prescriptive analytics and the term may just be a marketing ploy. We nevertheless provide several examples of this kind of analytics, noting that they might also fall under other categories of analytics.

Cognitive analytics go far beyond programmed answers, providing a technological approach that recognizes deep patterns in data and sifts through significant amounts of data to suggest solutions.

Cognitive analytics draw on all sorts of structured (survey data) and unstructured data (Facebook, social media) and can handle conceptual and symbolic data rather than just data collected from sensors or scripted survey questions (Gaikwad, 2018). For example, IBM's Watson, one of the first cognitive computing systems, can "identify the link between a contaminated water borehole, an epidemic of cholera, and low levels of school attendance in a region" (Wayua, 2014). IBM now has a research lab in Nairobi, Kenya, for Project Lucy ("IBM Brings," n.d.). Project Lucy aims to use Watson and other cognitive technologies to draw on big data to develop viable solutions to challenges in agriculture and beyond. IBM believes agriculture will benefit from cognitive analytics in five ways (Irimia, 2016):

1. Applying machine learning abilities to sensor and drone data will help the IoT achieve its maximum potential.
2. Image recognition can provide a basis for a drone's ability to capture images, analyze them in near-real time, identify areas of concern, and take actions (for example, apply pesticide if needed).
3. Many farming operations could be done remotely, removing the need for large numbers of people to work the land.
4. Cognitive analytics could be used to determine how certain seeds will react to different soils, weather forecasts, and local conditions as a basis for recommending the best crop or seed options to maximize returns ("Libelium," 2017).
5. Chatbots—conversational virtual assistants—for farmers can assist farmers with answers to their questions, giving advice and recommendations for specific problems.

There are startups in this space as well. Ricult has developed a platform that leverages cognitive data analytics. They designed their own proprietary creditworthiness assessment to link farmers with affordable credit, high quality farm inputs, extension services, and visibility to buyers—thereby addressing the farmer ecosystem. (See also the case study in Section XIII.)

Another interesting start-up is Farm.ink (see box). It is tackling the use of social media to design chatbot technology to help farmers identify livestock disease outbreaks and get real-time information. Chatbots may also be used in the future by farmers to negotiate market prices or order inputs (Raguse, 2017).

The challenge that cognitive analytics, and any sophisticated analytics, face in the agriculture sector is that industry does not fully appreciate that no two environments are alike—making testing, validating, and rolling these technologies out much more laborious in agriculture than most other industries. For cognitive solutions "to truly make an impact in the field, more effort, skills, and funding is needed to test these technologies in farmers' fields" and at a pace that makes sense for the market (versus responding to pressures that investors place on product development and roll-out) (Byrum, 2017).

FARM.INK: A MINI-CASE

Some of the largest Facebook groups focused on farming are located in Sub-Saharan Africa and include upwards of 100,000 people ("Africa," 2017). These groups discuss a vast range of topics including dairy farming and livestock. Thousands of historic posts detail exchanges on livestock diseases and queries on productivity and management. Until now, this data has not been leveraged to provide insights that could be helpful to farmers beyond their direct involvement in the farming groups.

Farm.ink, in collaboration with the International Livestock Research Institute (ILRI) and with support of CGIAR's Big Data Inspire Challenge, aims to combine social media data with ILRI data to create an open source platform that analyzes and visualizes emerging livestock disease outbreaks in Kenya. Farm.ink has developed a chatbot application in partnership with IDEO.org to communicate with farmers directly regarding farming concerns. They hope to also warn farmers of disease outbreaks and provide other information. Farm.ink's chatbot and digital farmer group products now cover over 100,000 users. Their chatbot product has produced user retention rates that far exceed those of smartphone apps and SMS services and send and receive millions of messages a month.

Georgia Barrie of Farm.ink shared insights into Farm.ink's cognitive analytics work (Barrie, 2018). Farm.ink's code uses natural language processing to tag posts made on Facebook by farmers, such as tags on the type of crop (e.g., "tomato") or location (e.g., Meru, Kenya). Farmers can then use the chatbot to filter posts by type of crop or type questions into the bot to find specific answers (such as one might use Quora or Stack Overflow). If a farmer types a question into the bot, the bot will use a search engine to find social posts and answers that most closely match the question.

Farm.ink collects some data directly from farmers, including types of crops they farm, whether or not they are currently farming, and their farm location. GPS is not often used since many people are not on the farm when using the application; mobile signals are still a challenge in rural areas so people tend to use the service when they are at home or where connections are stronger.

Using aggregate Facebook profile data that is automatically collected from users, Farm.ink knows what proportion of their users are male / female, their age brackets, and other data points like the types of handsets they use. This enables Farm.ink to generate rich profiles much faster than traditional SMS-type services, and their rates of registering this information are much higher than services based on basic phone technologies. By looking at the digital trace of farmers using their chatbot, Farm.ink learns a lot about their users and can tailor the chatbot experience on an individual customer level. Content can easily be tailored based on what crops they have stated and demonstrated interest in and on contacts with other farmers or buyers who are in their location. With the new collaboration with ILRI, Farm.ink aims to analyze trends in livestock health and identify disease outbreaks in real time.

Source: Interview with Georgia Barrie of Farm.ink

Table 3 below draws on Nordin's and Vlamis' constructs (Vlamis, 2015) for the different types of analytics we discussed with our key informants. It includes the different techniques these service providers said they utilize to analyze data about farms and farmers. (Any one entity may draw on several types of analytics to driving decision making.)

Table 3: Data Analytics Matrix supporting smallholder farmers

Time	Types of Analysis	Aims to answer the question…	Data sources/Requirements	Examples
Looks to the past	Descriptive analytics	What happened, when, and where?	Takes form of real-time dashboards, basic statistical analysis	CGAP smallholder farmer diaries data reports that outline key characteristics of farmers
	Diagnostic analytics	Why did it happen?	Takes form of analytic dashboards; use of decision trees, classification algorithms, and regression statistics	Use of Random Forest Algorithm in the CGAP diaries to classify farmers into different categories to understand different attributes and drivers of behavior; Grameen/TaroWorks tools that segment key outcomes data associated with poverty, gender, or GAP adoption rates
Looks to the future	Predictive analytics	What is likely to happen?	Uses historical data to explain data and make predictions. Needs computing software and tools and can require large volumes of data and computing power, including big data and its use of deep learning from various types of algorithms (such as neural network algorithms, natural language processing, complex event processing, etc.)	One Acre Fund, Gro Intelligence software; alternative credit scores
	Prescriptive analytics	What should be done? What can we do to make ___happen?	Relies on similar techniques as predictive analytics but goes a step further and uses analytics to prescribe a solution	Using aWhere, historical weather data, to decide whether to plant or not; MUIIS, Mahindi Master Game
	Cognitive analytics	What is the best action?	Uses similar techniques as predictive and prescriptive analytics, but the more data that cognitive analytics are exposed to, the more they learn; it is less about giving a computer parameters for decisions than about teaching it how to make decisions. Technology that uses cognitive analytics would be chatbots, automated customer support services.	Chatbots, facial/plant recognition. (e.g., Farm.ink agriculture chatbot)

Source: Nordin, 2016 and Vlamis, 2015 and author's own examples.

VIII. QUALITATIVE ANALYSIS AND INSIGHTS

Thick data, such as qualitative research gathered for human-centered design, and big data are both valuable for understanding smallholder farmers.

HCD holds much in common with the well-developed management science of target marketing for commercially designed and delivered products and services. HCD research often creates farmer archetypes or "personas" encompassing the typical needs, attitudes, and psychographic factors of different audience segments. A persona is a fictional, yet realistic, description of a typical or targeted user of a product. "A persona is an archetype instead of an actual living human, but personas should be described as if they were real people" (Harley, 2015).

Big data are quantitative. HCD and other qualitative data are sometimes called thick data (Wang, 2016). Some consensus is emerging that a hybrid approach that combines big data analytics with thick data is most useful in helping interpret the customer experience (Konduru, 2017). User stories or personas, often developed through qualitative methods, contain "emotion, something that no scrubbed and normalized dataset can deliver" (Konduru, 2017). While big data may be positioned to answer "what happened or will happen?" thick data can answer "why it happened or why it will happen." Thick data is more likely to provide insights into cultural barriers to using a given product or service or the likely fears that farmers might have in adopting a new practice. The development of user narratives or personas can help create compelling stories and humanize data that people will listen to when designing products or services or calling people to action (Martin, 2015).

Summary: Qualitative analysis and insights

- Thick data (such as data gathered via qualitative research for HCD of products and services) and big data are both critical.
- The use of "thick" and big data is fragmented and data created during the design phase of a project may not be leveraged for product uptake and for understanding impact.
- While there is great promise with today's buzzwords in *big data,* development of farmer personas or other qualitative ethnographic methods are needed to personalize quantitative analytics.

Key Considerations

- Does your organization use thick data to develop qualitative archetypes of farmers to understand typical farmers or users of your services? Could you currently do this or would this be a new process?
- How does your organization leverage both qualitative and quantitative data to understand typical and atypical farmer profiles?
- Are there opportunities to better integrate the use of thick data with quantitative data being collected and used?

While it seems that thick farmer persona data is often used at the product design stage, it is rarely used for evaluating product use after the initial design process has been completed. For example, if one archetype identified during the HCD research is described as a "high-tech farmer," are monitoring systems set up to predict or classify real users as "high-tech farmers" and then monitor their behavior based on expectations of this archetype? PHB Development CEO Philippe Breul shared that despite the effort to develop personas and conduct human-centered product design, he has yet to see service providers translate these personas into customer segments for ongoing product monitoring. MNOs are inclined to simply stick to more traditional user segments (men vs. women, urban user vs. rural user, etc.) within their existing datasets (Breul & Chassin, 2018). This can result in missed opportunities, at, and client loss, at worst.

Tricia Wang, in her piece, "Why Big Data Needs Thick Data," cited Nokia—at one point the world's largest cell phone company—as an example (Wang, 2016). She explained how Nokia undervalued its qualitative, "thick data" insights, contributing to its downfall and eventual sale to Microsoft. Wang found that Nokia over-relied on its numeric data and failed to act on qualitative insights revealing that its low-income customers were ready to pay for more expensive smartphones. This likely resulted in them losing significant market share.

Service providers should seek tighter collaborations between HCD processes and big data analytics.

There is an opportunity for stronger linkages between HCD research and the data analytics that can drive the day-to-day user experience. Organizations could capture quantitative data that specifically ties to the farmer archetypes developed at the design stage. For example, if "high tech farmers" during the HCD research phase tended to be male, living in an urban area, and well-educated, these indicators could be collected in the farmer profile and then clustered to monitor this particular segment or user group. This would require tighter collaboration from the outset between the staff who will manipulate farmer data during implementation and those participating in product design.

Given the growing volume of digitally-generated data and the capability of products and services to become more individualized through insights gained from data analytics, there is much to be gained by connecting HCD processes with data analytics. Abdullah Saquib of Viamo Pakistan was interviewed for this report. He acknowledged this gap and noted that personas created for human-centered product design have been used most effectively *post-product development* when the primary service provider works as actively with the HCD team as with the data analytics team.

Creating a stronger link between thick and big data appears critical. Otherwise, "we may lose the empathy we need to connect with the human behind the farmer," according to Philippe Bruel of PHB Development, who was interviewed for this report. Lisa Chassin, also of PHB Development, added, "We have to understand aspirations, pain points, and put faces to the data."

Agriclture Project for Naatal Mbay in Senegal, Xaume Olleros

IX. DATA SHARING

Summary: Data sharing

- Cloud-based systems are making it possible for more than one firm to access real-time data collected from the same farmer via private data-sharing platforms. It is important to establish who will be the long-term owner, particularly within a consortium of technical assistance providers, researchers, and private sector service providers who have different long-term roles.
- Open data is data that is freely available for anyone to use and analyze in developing new learning products. Open data facilitates innovative creation of 'new' data and generation of new insights.
- Open data management platforms must address tensions between open data and consumer protection/ privacy laws and establish data standards to facilitate interoperability of data.
- Standardized indicators and survey questions are important, but in the digital age, providing a common framework for mapping data is also important.

Key Considerations

- What types of data-sharing platforms do you currently use (Google, Box, Dropbox, TaroWorks, etc.)?
- Does your organization have a data-sharing policy? Who is data shared with and how? Is data shared in real time or through periodic reports?
- Is your organization a contributor to or user of others' data? How might you contribute?
- What types of partnerships may be necessary to better leverage others' data? (You do not need to re-create the wheel.)
- What would be the advantages or disadvantages of making your data open?
- What type of data could create more value if it were open source?
- Does your organization have a consumer protection and privacy policy and related processes that ensure policy compliance?
- What risks have you identified for the collection and use of farmer data? What misuse of data might be possible that would harm the farmer as well as your organization? How can these be mitigated?

Data sharing is an important contributor to big data analytics, and technology is facilitating greater use of common sets of data by multiple actors. This section describes how private data-sharing platforms and open data-sharing platforms are being used by service providers and their collaborators. It outlines the challenges and risks of sharing farmer data and the efforts underway to facilitate data sharing.

Private data-sharing platforms and cloud-based systems make it possible for more than one service provider to access real-time data collected from the same farmer.

TaroWorks (a customer relationship database built off Salesforce), is a private data-sharing platform used by Oxfam America's Women in Small Enterprise (WISE) program. WISE manages multiple country offices, partners, and language needs by leveraging a partner community on Salesforce ("Women," 2016). Users of the platform have limitations on what they can see, edit, and delete, depending on their roles within the program. When a data-sharing system is designed, the key owner of the data must think about which roles need access to what data.

Grameen uses TaroWorks for its FarmerLink project in the Philippines (Chang, 2017). Grameen staff, its agri-business partner, and the financial services partner all have access to the same data source; however, each has a different dashboard with data relevant to them. Users of the platform do not have equal access to the data. Whenever a platform has multiple users, an important consideration is who owns the data—particularly once a project ends. For example, while Grameen might drive the development of the platform to facilitate collaboration among several service providers, once the project is over, one of the local entities needs to own the platform. If there is not a clear value proposition for all service providers and this is not decided upfront, a platform may provide only short-term value.

Other proprietary systems combine both data collection and management. These include Olam's OFIS and Accenture's Connected Crop Solutions for Smallholder Farmers ("Digital Agriculture," n.d.). Accenture typically sells its technologies to farmer aggregators such as farmer associations, agribusinesses, and the like. One pilot in the Indian state of Karnataka connected smallholder farmers to producers who are also connected to agents, nutrient suppliers, and insect control suppliers ("Accenture," 2015).

Open data-sharing and management platforms facilitate creation of innovative 'new' data and learning products.

Open data is data that is freely available to anyone to use, analyze, and develop new learning products for. CGIAR defines big data as "open, harmonized, interoperable, and integrated datasets from multiple domains aimed to accelerate agricultural research and data use in service of development goals" (CGIAR, 2017a). With this definition in mind, open data facilitates access to big data for service providers to leverage and use for new benefits.

The International Open Data Charter promotes six principles supporting open data ("Principles," 2015). Data should be:

1. **Open by default:** But to make this work, people must be confident a system will not compromise their right to privacy.
2. **Timely and comprehensive:** Data is only valuable if it is relevant and it should be made available in original, unmodified form.
3. **Accessible and usable:** Data that is machine readable and free of charge is accessible.
4. **Comparable and interoperable:** Data standards and ontologies play a crucial role in facilitating this.
5. **Used for improved governance and citizen engagement:** Transparency of data allows citizens to know what public officials and politicians are doing and hold governments accountable for improved public services.
6. **For inclusive development and innovation:** Open access to data can spur inclusive economic development—for governments and entrepreneurs alike.

The box below highlights how four service providers have provided open access to data, creating new value.

Data sharing and management platforms

CGIAR Big Data Platform in Agriculture

In late 2017, during the First Annual CGIAR Convention on Big Data in Agriculture, CGIAR revealed the prototype of a searchable CGIAR-wide data harvester (http://bigdata.cgiar.org/). CGIAR recognized that as many as 185,000 surveys are conducted each year within their own network of affiliated research institutes (Brian King, 2017). They needed a platform to organize their own data and make it discoverable, re-usable, and interoperable. (This platform is described in greater depth in the CIAT case study in Section XIII.)

i2i Data Portal

i2i, a global resource center that seeks to improve financial inclusion through thwe smarter use of data, was launched in 2015 and is jointly hosted by Cenfri and FinMark Trust. It is funded by the Bill & Melinda Gates Foundation in partnership with The MasterCard Foundation. i2i launched their i2i Data Portal (http://i2ifacility.org/data-portal) to share insights from the CGAP Smallholder Farmer Diaries and other national survey data. i2i intends to make additional relevant datasets and tools publicly available through this data portal.

Smallholder Finance Product Explorer

In 2017, the MIX Market, One Acre Fund, and the RAF Learning Lab collaborated on the design of a data framework to categorize a diverse set of smallholder finance products and allow comparison and benchmarking of financial products (https://www.themix.org/mixmarket/smallholderfinance). Currently in beta form, the Smallholder Finance Product Explorer provides information from nearly 30 financial services providers in ten countries. They plan to continue to add to the Explorer database as other smallholder finance providers share their data. This will enable greater learning by all database users, with the goal of reducing the financing gap for smallholder farmers.

Global Open Data for Agriculture and Nutrition (GODAN)

GODAN was launched in 2013 and has nine core partners: US Government, the UK Department for International Development, the Government of the Netherlands, FAO, CTA, Global Forum on Agriculture Research, The Open Data Institute, CGIAR, and the Centre for Agriculture and Biosciences International. Most efforts by GODAN have been to build high-level support for open data among governments, policymakers, international organizations, and businesses. A recent effort by GODAN and other actors has been to launch a beta version of the Agriculture Open Data Package (AgPack) which aims to help provide governments with a roadmap to publish agriculture data as open data (https://opendatacharter.net/agriculture-open-data-package). Most recently, USAID, DFID, and the Bill & Melinda Gates Foundation launched a joint effort (as funders of international agricultural research) to develop more harmonized approaches towards open data.

Open data management platforms must address tensions between open data and consumer protection/privacy laws.

Data privacy and consumer protection are important considerations for open data. The abuse of open data and big data analytics "could turn the 'Information Society' into the 'Surveillance Society'" (Biggs, Garrity, LaSalle, Polomska, & Pepper, 2016).

Much data on people is collected passively (from their mobile phones, usage data, etc.) but there are very few standard opt-in policies for data sharing (Caire et al., 2017). Data privacy laws vary from country to country. Countries like Ghana, South Africa, and Uganda stand out because their laws and regulations are guided by customer centricity principles such as:

- empowering the consumer to make decisions about their personal data usage, especially in relation to automated decision making
- stipulating clear mechanisms through which the consumer can seek compensation
- ensuring the customer has the "right to be forgotten"

Sharing of data across borders presents special considerations. Some countries stipulate that data can only be transferred to users in countries with laws providing the same (or higher) level of protection as their own.

An interesting model to consider is the American Farm Bureau Federation initiative to establish the Privacy and Security Principles for Farm Data ("The Privacy," 2018). These principles currently have 37 endorsers. They provide guidance on the processes that need to be considered to protect data privacy—such as collecting the minimum data needed in order to respect the farmer's time; obtaining informed consent; and being transparent about how data will be used. The model follows an approach similar to that of the Principles for Digital Development ("The Principles," 2018) and the Smart Campaign Client Protection Principles for the microfinance sector ("The Smart Campaign," 2018).

Standardized indicators and frameworks facilitate data sharing and data integration.

The two key themes for data standardization are: 1) *data/indicator* standardization or a common use of shared indicators and 2) *data ontologies* that help ensure data is interoperable and creates value for those beyond the direct beneficiary. Efforts such as those underway with the First Mile Reference Framework (referenced earlier) are contributing to the goal of standardization.

Andre Jellema from Data-Impact, interviewed for this report, found that **you need a flexible data framework that people can map their specific data to. This helps make data FAIR—Findable, Accessible, Interoperable, and Reusable** and ensures multiple service providers can access data from a common framework. Jellema shared his experience working with UTZ/Rainforest Alliance for the First Mile Data Framework and assisting in the development of the GODAN AgPack. He noted that creating standardized indicators and survey questions is one stream of effort. But in the digital age, providing a framework that multiple service providers can use to map their data is also important. The development of controlled vocabularies ("Use COAR," 2018) and harmonization of crop ontologies are very important to this process (The AIMS Team, 2016). He noted that unless you are a scientific expert in the field, understanding these ontologies might be difficult. Thus, a tiered approach may be useful for different types of users: developing more basic ontologies for non-expert users and more specific/comprehensive ones for the scientific community.

For the financial services sector, standardization of key indicators has allowed greater transparency and learning across countries and regions. Involved entities include the MIX Market, the Findex surveys, and Finscope. The IRIS initiative of the Global Impact Investing Network, a nonprofit organization dedicated to increasing the scale and effectiveness of impact investing, has worked to standardize indicators for those investing in agriculture projects ("IRIS," 2018). The USAID-funded FANTA project developed a guide in the 1990s on seven key agriculture indicators for food aid programs (Diskin, 1999), which has since been updated for Feed the Future (Nelson & Swindale, 2013). The FAO has also developed their own compendium of nutri-sensitive agriculture indicators (FAO, 2016), along with strategies to advocate and promote open access of their data (Caracciolo & Keizer, 2012). CGIAR has their own community of practice, led by Bioversity International, that develops ontologies as part of the CGIAR Platform for Big Data in Agriculture ("Ontologies," 2018).

X. DATA USE

Summary: Data use

- Service providers are using farmer profile data for similar purposes as in the past. But advances in technology are enabling better use of farmer profile data to improve processes, deepen understanding of farmers' needs, gain insights from multiple data sources, and provide more responsive support to farmers.
- Service providers—and farmers—should not treat data just as a resource, but as an asset.
- Data is giving rise to new business models.
- Data monetization can provide an evolving income flow to support business models.

Key Considerations

- How might farmers be compensated for sharing their data?
- What would you consider your data assets?
- What might other service providers consider as your data assets?
- Data is giving rise to a new economy. Is there an opportunity to monetize your data assets?

This section summarizes how service providers use data and how data can be better used in the future.

Changes in the use of farmer profile data is changing how we serve farmers.

Advances in technology are changing how farmer profile data is used to improve processes, deepen understanding of farmers' needs, gain insights from multiple data sources, provide more responsive support to farmers. These advances are taking place across the chain of data-driven activities that support smallholder farmers. They include:

- **Farmer identification:** Farmer profile data is used to engage farmers targeted to receive tailored products and services. With access to new technologies, commercial service providers are using data analytics to predict users of agricultural data, even when farmer profile data is lacking. For example, using the example provided earlier, MNOs can use limited phone usage data to predict whether the phone user is a man or woman.
- **Farm data:** Data related to the farm, such as plot location, size, soil quality, and crops grown, along with plant or animal science, drive decisions related to extension services, profit potential, and connections to markets and financial services. New technologies rely less on extension agent interactions with the farmer to collect data and more on remote technologies such as satellite, sensors, and drones.
- **Product and service design:** Along with qualitative data (thick data), quantitative farmer profile data and data analytics can be used to design and develop extension/knowledge services, financial services, and market access services. For example, connecting specific farmer profile data to satellite and weather data can highlight a farmer's need for advice in responding to a crop disease outbreak that might not otherwise have been detected.
- **Direct support to farmers:** Farmer profile data is used to provide services (information, financial services, market support) through person-to-person approaches (e.g., extension agent services), mobile-to-person approaches (e.g., SMS messages), or farmer apps (e.g., Farm.ink, iShamba). Extension agents are often responsible for many more farmers than they can visit in a timely manner. Analysis of farmer profile data can help predict and prescribe solutions in real time and help provide immediate support to farmers. Furthermore, these digital services can target hard-to-reach farmers.

- **Monitoring and uptake:** Farmer profile data can be used to track changes in product/service usage and to monitor farmer behaviors and outcomes (e.g., adoption of GAPs, yield changes). Use of traditional paper-based farmer profile data can allow for gaps in data sharing between research team members and program staff. When farmer profile data is in the cloud and tracked over time by multiple service providers, it allows immediate access to farmer behaviors and outcomes.
- **Data sharing:** Three types of sharing are common among service providers. These include private data sharing among multiple stakeholders (e.g., project-based data sharing among two-to-four service providers); open data sharing where data is public and available for any service provider to use (e.g., CIAT's Big Data Platform, described below); and business intelligence services typically driven by commercial entities that purchase and sell data (e.g., Gro Intelligence).
- **Research and evaluation:** Farmer profile data has been traditionally used to conduct impact assessments or carry out research regarding farmer support services (e.g., CGAP smallholder diaries). Data analytics and increased open-data sharing policies and practices are advancing and changing how research and evaluation are conducted. No longer are researchers and program teams relying on single datasets. Rather, multiple datasets are combined to deepen and broaden service providers' understanding of program impact.

These points are summarized in Figure 5, on page 44, which highlights the difference between traditional and new sources of data and how they support farmer outcomes.

Service providers—and farmers—should not treat data as just a resource, but as an asset.

Jerry Glover, from the USAID Office of Agriculture, Research, and Policy within the Bureau of Food Security, noted that we spend a great deal of time and money collecting data on farmers but not much on managing that data. We should consider **how to treat data not only as a resource, but also as an asset.** Michael Shrage, a contributor to Harvard Business Review, estimated in 2016 that in the next 12 to 18 months, service providers could expect to have access to 10 to 1,000 times more data (Schrage, 2016). He said this data provides opportunities to create new value. "Data oversight is the next leadership challenge for organizations." According to Jason Tatge, CEO and co-founder of Farmobile (who was interviewed for this report):

> *"...data is the new currency. If farmers are going to adopt precision agriculture and other modern farm technologies, we cannot afford to ignore the value inherent in the data collected. Farmers who collect and own their data are natural partners with companies who are trying to validate their product performance. The key is to properly align the incentive structure. Farmers need a reason to ensure that good data is collected."*

Farmers should treat their data as their property, not just information shared with others. **"Data has value,"** Dan Berne shared (Berne, 2018). **"If you treat your data like property, it tends to be more yours, like physical property."** While the industry is still struggling with how to compensate farmers for their data, this begins with a mind shift—from treating data as a resource to treating it as an asset or property that has value and deserves protection.

Data is giving rise to new configurations within service provider models.

Particularly among B2B and B2C service providers and those promoting open data/big data sharing platforms, the collection, use and analysis of farmer profile data are part of business models that are meant to result in financial sustainability or in public goods. Service providers like aWhere and Gro Intelligence are selling data. CIAT and CGIAR are developing open data-sharing platforms. Elsewhere it is being said that **data is giving rise to a new economy** ("Data," 2017). **Data can be monetized as an asset.** This is considered by Gartner, Inc. as "infonomics," or the assertion that information has economic value and significance (Laney, 2018). Information should not only be spoken or talked about as an asset but also treated as one. Some service providers interviewed here are starting to do that. Even open data sharing, which assumes free and transparent data, can potentially be monetized when data services such as data aggregation or data analytics services are sold that help others make efficient use of existing open data or helping make their data open.

In a 2016 Alliance for a Green Revolution in Africa (AGRA) report called *Digital Harvest,* Advantech Consulting Ltd. noted that out of 15 ICT4Ag solution providers across Kenya, Tanzania, and Ghana, only five were delivering services sustainably (Waruingi & Muriithi, 2016). Those who were successful had hybrid business models with data monetization as one means to achieve sustainability. Examples included service providers (or governments) paying for smallholders to access services (payer is different from the user); providing both digital and face-to-face interactions; offering a combination of services; and offering copyrighted technology and deriving income from diverse sources (subscription and usage fees, data monetization, bundling of services that farmers were willing to pay for—such as credit—with extension services). It is unclear at the time of this writing whether data monetization will become an important revenue stream for most of service models.

For some actors like MNOs, there must be a strong business case to provide agriculture value-added services (Agri VAS). For the GSMA mAgri pilots, the average investment was upwards of $700,000 (Palmer & Darabian, 2017b). Some of the MNOs found that B2B revenue models could subsidize the costs of the end-user (e.g., by selling MNO data to a fertilizer company so that they could use this data to identify potential customers). Four of the six MNOs in the pilots had increased average revenue per user among Agri VAS users compared to non-users. Some MNOs charged small amounts to the user. For example, Dialog Sri Lanka charged 1 LKR (less than 0.007 USD) per day per crop.

Service providers like Ricult, MUIIS, and Akorion serve as B2C aggregators. This means they directly charge fees for services (whether for accessing information, financial services, input supply payments, etc.) but also ensure their farmers are provided directly or linked to complementary services within the farmer's ecosystem (for example, they are linked to input suppliers and buyers). B2C aggregators may also share revenue or profit directly with other service providers, such as when one provider sells its customer data to others (such as the MNO example above). These B2C aggregators aggregate farmer data and provide bundled services that essentially cross-subsidize each other (credit fuels free information services, for example).

Figure 6 below illustrates various data flows and data monetization flows. On the left side of the graphic are service providers who primarily rely on donor or government funding (such as researchers, technical assistance providers, NGOs, and government extension service providers). On the right side are commercial entities such as B2B and B2C service providers. In the center are the farmers who are the primary sources of data and beneficiaries of farmer-support services. Farmers provide data to most service providers in exchange for support services. While the earlier discussion on blockchain suggests that in the future farmers might monetize their own data, this is not shown here since it apparently is not yet occurring.

Donors and investors continue to support both the NGO/government and commercial sides of the equation. The expectation is that over time, less support will go to the commercial side. For the foreseeable future, however, donors and investors will continue to support the B2C models, especially as they are being tested for sustainability (e.g., the models of Ricult, MUIIS, and Akorion).

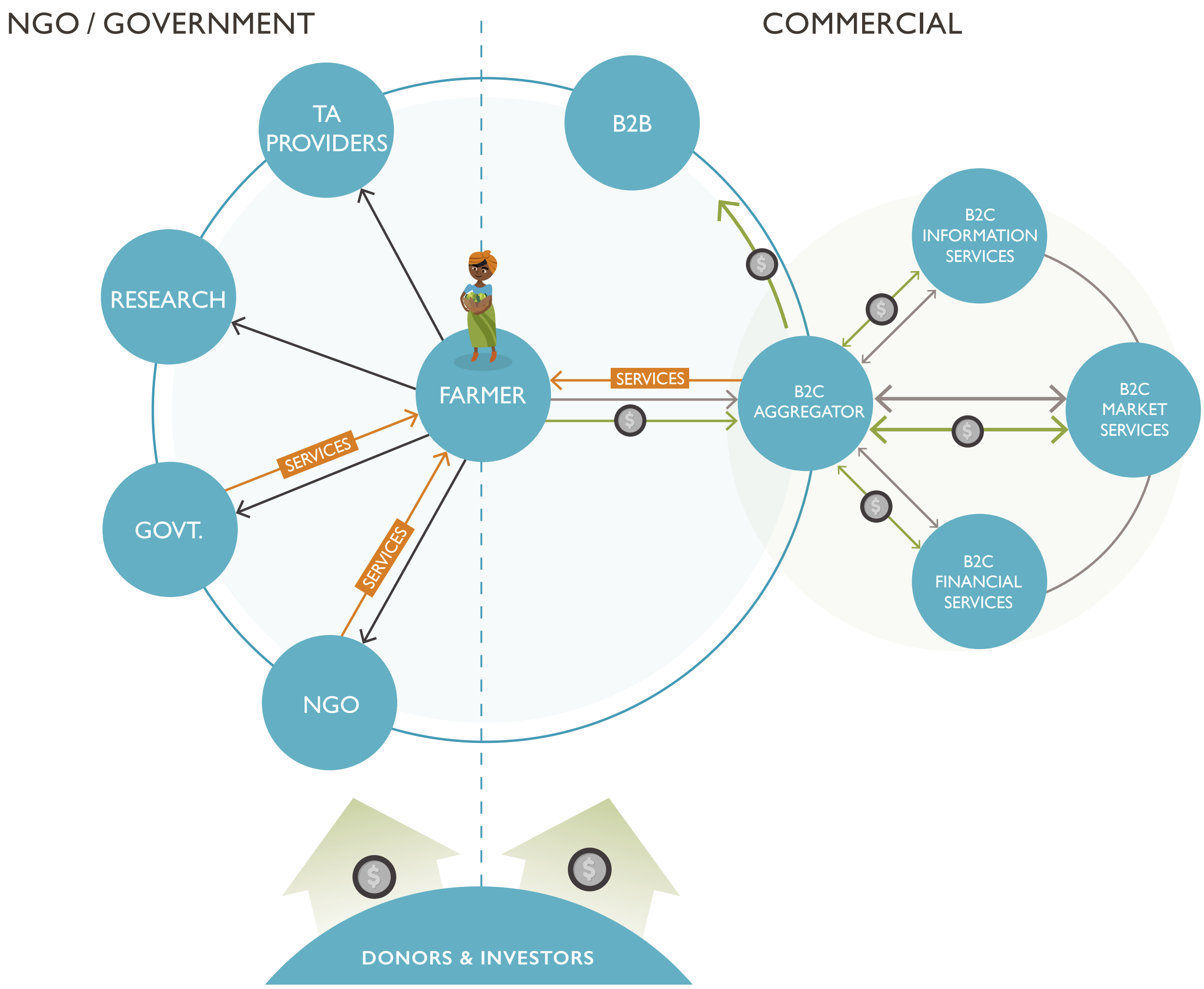
Like the B2C aggregators depicted in the graphic, NGOs or technical service providers can also function as aggregators.

Figure 5: Leveraging new technologies to strengthen farmer-support activities

	Farmer and Needs Identification			Direct Support to Farmers			Evaluation and Sharing		
	Farmer Identification	Farm Data	Product & Service Design	Extension Services/ Information	Financial Services	Market Access	Monitoring and Uptake	Data Sharing	Research & Evaluation
New Data Sources	API to identity data Biometric data Phone usage	GPS Satellite Drones Sensors Crowd-sourcing	Analytics that provide insights on farmers and farms Social media and other unstructured data	SMS IVR/OBD Apps Chatbots	Realtime transactions: with APIs to financial institutions & apps Blockchain Alternative credit scores Willingness to pay metrics	Apps that collect & provide real-time pricing data, sales, costs	App profiles IVR/OBD Social media Real-time tracking of progress	Private and open-source data sharing platforms	Evaluations that can draw on other open-source data and leverage data analytics to provide new insights
Traditional Data Sources	Farmer's associations Farmer groups Women's groups	Physical field mapping Soil sampling Field extension agents In-person surveys collecting snapshot data	HCD research & other in-person research	In-person Extension services In-person surveys collecting snapshot data	In-person credit officer services In-person surveys collecting snapshot data Group lending practices	In-person surveys collecting snapshot data Extension services	Field extension agents In-person surveys collecting snapshot data	Partner data shared through reports, in person meetings	Single data-sets related to project/singular effort

For example, organizations like Mercy Corps, Grameen Foundation, and CIAT can function as network orchestrators that bring B2B, B2C, researchers, and government actors together in service of developing a farmer ecosystem. There can also be exchanges of money between any of the actors; for example, from researchers to B2B actors. (However, for simplicity's sake, these examples have been omitted.)

Figure 6: Data and revenue flows



XI. WHAT IS INNOVATION IN FARMER PROFILE DATA MANAGEMENT

Summary: Innovation in farmer profile management

- How service providers use and manage data, in combination with the methods and data analytics they use, determines how innovative their model may be.
- There is no single pathway to sophisticated use of farmer profile data; even small data can be big data depending on the sources of data used.
- Consider emerging technologies and assess whether you should make an early move, take a moderate path, or wait for further maturation of the technologies and their uses.
- To avoid being overwhelmed, take baby steps.

Key Considerations

- What current combinations of data points, data capture methods, analytics, and data use and management does your organization use? What new combinations might you consider?
- How can you utilize newer data capture methods or analytics to improve your farmer profile data?
- What energy does management have to transform the organization so that the data management practices yield better decisions?
- Start small and lean. With many options to consider, prioritize those that align best with your strategy and provide quick wins and momentum.
- How comfortable is your organization in taking risks? What benefits might there be to being an early adopter of new technologies? What might be the risks?
- For which technologies do you want to see more use cases?
- How might your organization develop a new use case for a data management approach?
- Will utilizing established and known technologies accomplish your goals?

Data monetization is only one means to achieve sustainability of farmer support services and while data monetization may not be possible for all service providers, it is an untapped opportunity for some.
This section summarizes how service providers can assess their own data management practices and consider use of the new technologies available.

Even small data can be big data in the right combinations. Service provider models are all experiencing innovations, but individual stakeholders are at different points in their uses of farmer profile data. **Even small, simple data can be big data in the right combinations.** Not every service provider may want or be able to analyze their farmer profiles in the context of big data—but this need not prevent them from contributing to the big data economy.

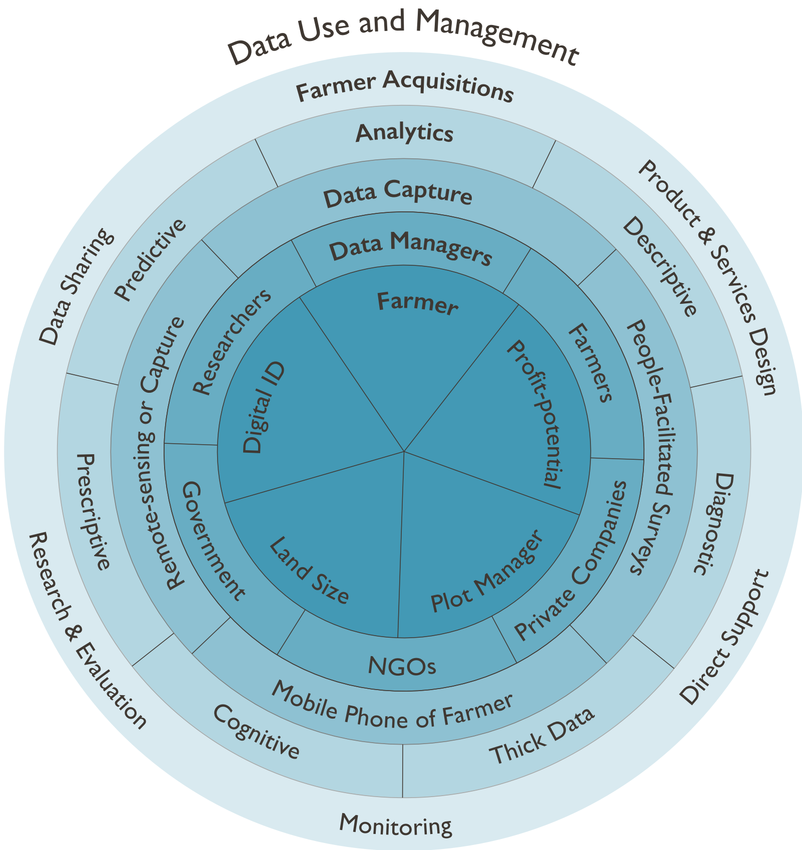
There is no single pathway to sophisticated and innovative use of farmer profile data.

Figure 7 below demonstrates this point. **Sophisticated and innovative use of farmer profile data depends on how service providers manage the data, deploy data analytics, and use the data.**

Imagine a calculator wheel that looks like the figure and that each concentric circle could spin by itself. If you aligned different combinations of the wheel, you could determine whether the service provider was using digital data assets to their fullest. For example, this report has provided potentially ground-breaking examples of combining farmer data with weather and sensing data; using predictive analytics; and making this data a common good.

In contrast, well-worn practices such as collecting data for an impact assessment, and conducting only descriptive analytics with limited distribution and sharing, generally fail to create new value for those beyond the immediate stakeholders. However, if this same data set would be opened to others or combined with new data forms such as satellite data, it could create new value.

Figure 7: Farmer profile data wheel



The wheel can be used as a tool for reflecting on current service provider models and project new combinations for better or more innovative use of farmer profile data. Just as there is not one pathway to sophisticated use of farmer profile data, neither is there a necessity to do everything.

Consider emerging technologies and assess comfort with risk.

For a service provider who wants to consider new, innovative combinations, Gartner Hype Cycles help assess the promise of emerging technologies along with a service provider's appetite for risk (see box). Gartner's Hype Cycle is a graphical depiction of a common pattern that arises with each new technology or other innovation. Each year, Gartner creates more than 90 Hype Cycles in various domains as a way for clients to track technology maturity and future potential. The cycles are useful in helping decide whether to make an early move (combining risk of failure with the potential to reap the rewards of being a first mover), take a moderate approach (insisting on sound cost-benefit analyses when new things are not yet fully proven), or wait for further maturation (wait until others can deliver tangible value).

The definitions of the Hype Cycle phases are provided in the box below.

Hype Cycle Definitions

Each Hype Cycle drills down into the five key phases of a technology's life cycle.

Innovation trigger: A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger significant publicity. Often no usable products exist and commercial viability is unproven.

Peak of inflated expectations: Early publicity produces many success stories—often accompanied by scores of failures. Some companies act; many do not.

Trough of disillusionment: Interest wanes as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters.

Slope of enlightenment: More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious.

Plateau of productivity: Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

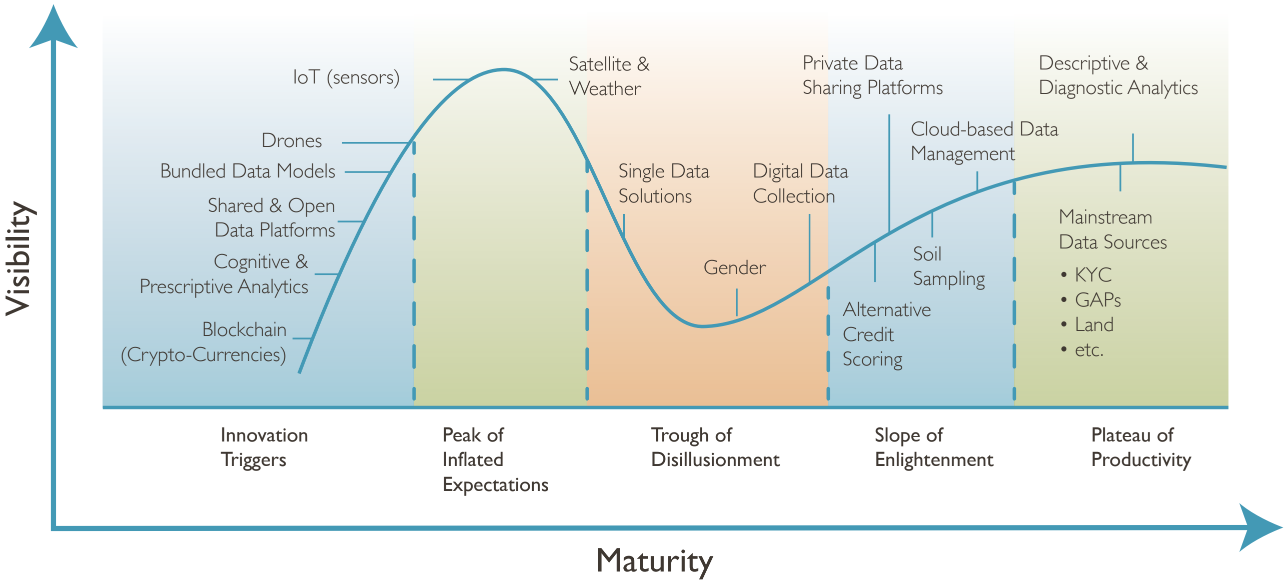
Source: Gartner Hype Cycle. (2018). Retrieved from https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

Ryan Rakestraw, of Monsanto Growth Ventures, applied the Gartner Hype Cycle methodology to help his team evaluate the maturity level of different agricultural technologies and implications for their own investments (Trice, 2016). In Figure 8, we adapted Rakestraw's Gartner Hype Cycle for agricultural technologies (Rakestraw, 2016) to the context of smallholder farmers, looking specifically at farmer data sources and management.

In the figure 8, data (collected from or on the farmer/farm) and data management approaches (including data analytics, sharing, and governance) are mapped on a curve illustrating both visibility and maturity of innovations. (The placement of different data sources and models are for illustrative purposes only; their placement can be debated.) As in Rakestraw's Monsanto curve, many data sources mapped here are simply innovations in the agriculture-technology space (e.g., providing weather data to farmers as a service versus using weather data to create new data for analytics).

Although Rakestraw's Monsanto curve likely represents a different target market, it is interesting to note that some data sources and approaches mapped on Rakestraw's Monsanto's curve are also found on the one for smallholder farmers. (For example, blockchain technology, despite being around for at least a decade, is placed as an Innovation Trigger on Monsanto's curve and also for smallholder farmer data profiles.)

Figure 8: Farmer profile data hype curve



The following bullets explain placement of the data or data management approaches on the hype curve:

- **Innovation triggers:** Blockchain, shared and open data platforms, bundled data models, cognitive and prescriptive analytics, and drones are being discussed but are relatively new to the market and have to prove their viability and practicality.
- **Peak of inflated expectations:** IoT (sensors) and satellite and weather data use is accelerating. But challenges include difficulty in interpreting data by non-experts, lack of consistent power sources and connectivity, calibration and cost of sensors, training in how to use sensors, network bandwidth challenges as more devices are connected to the internet, and cyber security threats.
- **Trough of disillusionment:** Business-to-customer single data solutions face sustainability issues because they do not address the comprehensive needs of a farmer and may have limited value to the farmer. Gender equality remains a challenge generally for agriculture, but it also remains a challenge because the sector lacks good data on women.
- **Slope of enlightenment**: Digital data collection, cloud-based data management, private data-sharing platforms, alternative credit scoring, and soil sampling fall into this category because they are becoming more common and business cases for their use are clearer.
- **Plateau of productivity:** Data sources like paper-based surveys are mainstream even as they have continued to evolve. Descriptive and diagnostic analytics are commonly used, particularly among service providers with skilled analysts.

Start with small, manageable steps.

Successful data-driven agriculture need not incorporate everything mentioned in this report. It is important to **start small, with manageable steps.** Dan Berne recommends: (Berne, 2018):

- Start with the end in mind. (What does the farmer want to achieve—increased yield? profitability? What is it you are trying to help farmers improve?)
- Take baby steps. There is so much you can do with data. Pick a goal and identify some easy wins. You can quickly get overwhelmed.
- Be patient. It may take a few years to see the full value of your efforts.
- Get agreement on what data you will provide and how it will be used. Make sure both the farmer and service provider are in agreement on what data will be provided and how it will be used.
- Make sure you have support and learn from others.

XII. CONCLUSION & NEXT STEPS

Summary

- Transformational use of big data could bring fragmented data, resources, and service providers together to build a more supportive farmer ecosystem.
- Farmers need to make decisions at *crucial moments*; the aggregation of their profile data, remote-sensing data, satellite data, and weather data makes this possible.
- Now is the right time to create ecosystems that connect farmers with a multi-disciplinary set of actors committed to engaging and doing business in rural communities because:
  - We can clarify the farmer data we have and get the data we need.
  - We can better marry plant/animal science with human science.
  - We have an opportunity to fully link the farmer with the value chain.
  - We have an opportunity to be precise.
- Despite the growth of the digital economy, person-to-person support will still be necessary to build trust and bring farmers into the digital ecosystem.

Robert Cailliau, who co-designed the World Wide Web, stated, "When we have all data online it will be great for humanity. It is a prerequisite to solving many problems that humankind faces" (Silva, 2009). Never in the history of mankind have we had access to so much information and so easily, but we need to better manage and make use of that data. This final section summarizes the key findings and concludes with thoughts regarding the development of an ecosystem that fully leverages farmer profile data and supports farmers with timely and necessary services.

Smallholder farmer data management must begin by defining from whom data is captured, how it is captured, how it is analyzed, and how it is used and shared. Key findings include:

- Defining the smallholder farmer is not easy; the group is not homogeneous. The definition of a smallholder farmer must be flexible to avoid excluding particularly vulnerable people.
- Given there can be several farmers per plot of land; when aggregating data for either open data efforts or for sophisticated analytics, it is the farm itself that pulls data together.
- Service providers who capture data from, about, and for farmers are a diverse group of stakeholders. The type of service provider does not necessarily determine what they collect or how data is used, but it is an important starting point.
- New ways of collecting and aggregating data and applying analytics—such as predictive, prescriptive, and cognitive analytics—can reduce the amount of direct input needed from the farmer. Data analytics are a game-changer and are being used to create "new data" from existing data.
- Technology is facilitating the sharing and management of farmer profile data in real time.
- Marrying plant science data with real-time farmer data is a new frontier for improving farm productivity.
- Service providers are winning or losing depending on how they use their data. Service providers and farmers should not treat data only as a resource but also as an asset; they should consider opportunities for data monetization. Smallholder farmer data is giving rise to new business models.
- There is no single pathway to sophisticated use of farmer profile data. Sophisticated and innovative use of farmer profile data depends on how service providers manage the data, deploy data analytics, and use the data.
- When considering how to integrate new farmer profile approaches, start small and take manageable steps.

Transformational use of "big data" could bring fragmented data, resources, and service providers together to build a more supportive farmer ecosystem.

This assessment revealed that most of the data and the technology (hardware and software) already exist to solve many constraints faced by farmers. But they are fragmented and not all service providers—or farmers—have equal opportunities to access them. Use of big data could bring fragmented data and resources and diverse service providers together for a more supportive farmer ecosystem. It can help provide smallholder farmers with financial services, market access, input supplies, and information services at crucial junctures. The first step is knowing who the farmer is and what his or her needs are. Digitizing data from and for different stakeholders within the farmer's ecosystem can improve transparency and trust among the farmers and service providers and integrate the farmer more effectively within the overall value chain to improve productivity, profitability, and the well-being of smallholder farmers and their households.

Christian Merz, who who previous worked for the for the Agricultural Development Program of the Bill & Melinda Gates Foundation as the Senior Program Officer for Digital Solutions, highlighted that **it is time to move beyond single solutions and put more energy into developing the ecosystem** (Merz, 2018). Donors, in his opinion, are still necessary to support systemic innovation. However, instead of funding "individual winners," the Gates Foundation wants to finance initiatives that strengthen multiple interacting actors. An example is the Africa Soil Information Service, a Gates-funded initiative that aggregates and packages data on soil health.

Advances in technology open unprecedented opportunities to strengthen farmer ecosystems.

Advances in technology and digital infrastructure make this an important time to strengthen linkages among multi-disciplinary actors committed to engaging with and doing business in rural communities for the benefit of smallholder farmers.

1. **We have the opportunity to clarify the farmer data we have and get the data we need.** How many of the same farmers are asked the same questions over and over by different service providers and how many are left out? How many women or other vulnerable groups are left out? It is hard to design for and service the needs of people if you have no information about them. Blockchain and mobile phone ownership may give visibility to people who need to be included. Practical data management barriers prevent true open data sharing. Until these are solved at both the macro and micro level, disjointed data collection efforts resulting in stovepipe repositories and farmer fatigue will unfortunately remain the norm.
2. **We have an opportunity to marry plant/animal science with human science.** Plant and animal science research is crucial to increase yields and improve animal health. Organizations such as CGIAR have abundant scientific insights relevant to smallholder farmers. In addition, **the aggregation of information from a farmer's profile, remote-sensing data, satellite data and weather data makes it possible for farmers to make decisions at *critical moments*.** Data analytics that pull data from multiple services can be used to provide timely, individualized information based on the farmer's and the environment's current and predicted conditions.
3. **We have an opportunity to link the farmer with the full agricultural ecosystem**. Markets themselves are underdeveloped. Access to agricultural financing has been historically limited. While lack of data is not the only reason for poor markets or limited access to financing, it is an important one. Buyers do not have information on smallholder farmers and neither do many financial institutions. Data on farmers' harvests and productivity creates this visibility for other stakeholders in the value chain; data analytics can also directly link people to markets and financial services.
4. **We have an opportunity to be precise.** Many service providers and projects that have begun to integrate technology writ large have been able to show cost-efficiencies in their service delivery and demonstrate impact. If a farmer had a better market, an agricultural loan, and an agriculture extension agent that showed up every day, would that improve his/her yield? Likely so, but would this be sufficient in the face of environmental, physical changes? Precision has been missing. **More is required than information about the farmer and about predictable behaviors; "smart information" is needed to help inform choices during unpredictable circumstances.** Farmers need to understand the tradeoffs of

different financial decisions such as "If I use this amount of fertilizer that costs me this much, what will my yield be? Which market has the best price today?" **Farm development plans and other tools that model these circumstances help farmers make more informed decisions based on their own financial situations.** Such precision may better support a smallholder farmer who has no room for financial error; he or she may find it worth the money to pay for such services.

In tomorrow's data-driven farmer ecosystem, service providers will help orchestrate activities among complementary actors.

Some of the most interesting models explored in this report are of service providers who pull together many of the opportunities mentioned above. They use APIs to combine multiple sources of data—their own and that of others—to provide both in-person and automated recommendations. They provide credit directly or provide insurance to farmers and help them communicate and negotiate with buyers. They charge farmers or rely on financial services within a program to cross-subsidize costs, or they monetize their data for use by other service providers. Testing the impact of these approaches will be important.

Person-to-person support will still be necessary to build trust and bring farmers into the digital ecosystem.

Despite the opportunities of big data, precision agriculture, data analytics, and data monetization, well-known challenges lie ahead. Data privacy is one. Building the capacity of stakeholders is another. Developing digital strategies and data science expertise require investment. Person-to-person support will still be necessary to build trust and bring farmers into the digital ecosystem. The right big data infrastructure is crucial for success, but human and cultural factors can be critical. "People are the big data difference makers" says Teradata, a big data analytics firm ("Building a Team," 2018). Usman Javaid from Ricult highlights this from his experience:

"There is a big trust deficit. Any element that is not indigenous from where [the farmer] comes from, that trust deficit is due to past scams and exploitation. Farmers do not trust people easily. Trust grows but it is a slow process. [Farmers] need to believe in the technology themselves. It is a question of time."

He believes that once farmers trust a service, they can interact with it and require less personal time with an agent.

The pieces are in place to support data-driven agriculture; the question is how to best bring them together for the benefit of the smallholder farmer.

Many questions remain. As summarized in Annex 2, some questions can be answered by service providers themselves. They can determine how to leverage their data assets effectively and explore the roles they could or should play in ecosystem development. But as this landscape assessment has found, all the pieces are here to support data-driven agriculture. The question is how best to bring those pieces together for the benefit of the farmer and the world he or she is expected to feed.

XIII. CASE STUDIES ON INNOVATIVE FARMER PROFILE DATA MANAGEMENT

The work of three service providers highlights innovative approaches to the development and use of smallholder farmer profile data and support of the smallholder ecosystem. In this section we include three case studies:

- **Ricult is a private company** that provides comprehensive support services to farmers and leverages many of the modern data technologies available for service providers working in agriculture.
- **CIAT/CGIAR and their Big Data Platform for Agriculture** shows how researchers can leverage publicly available research and evaluation datasets to contribute to new learning in the agriculture sector.
- **Grameen Foundation's ICT4Ag approach** highlights how technical assistance service providers engage in partnerships to develop data management systems that support extension activities implemented by public, private, and non-governmental organizations.

Table 4 below shows key features of the three case studies.

Table 4: Main features of farmer profile data management in three specific service provider models

	CIAT/CGIAR	Grameen Foundation	Ricult
Service provider model	**Research**	**Technical Assistance/Project Management**	**Commercial**
Theme of case study	Open data; big data platform, big data analytics for agriculture	Integration of new data management tools and processes in relationship with partners and through project-based efforts	Developing a business model that responds to farmers' needs for financial services, agriculture information and support, market access
Types of data sources	CGIAR network members and their thousands of datasets openly available on a searchable platform; APIs and access to other data sources such as satellite, weather, partner data, etc.	Databases shared on TaroWorks and other platforms by partners during project; primary focus is on farmer adoption of GAPs, access/use of financial services, certification, farmer development plans	Proprietary database of farmer information, satellite data, APIs to other databases of information
Data capture methodologies (primary data)	Varies based on member organizations and their projects; can include digital data collection, traditional surveys for research/evaluation	Primarily digitally- enabled, people-facilitated data capture; TaroWorks and other mobile data collection tools such as SurveyToGo Community Knowledge Workers and field agents collect data using mobile devices; traditional surveys also support research and evaluation	Both people-facilitated and mobile phone-data capture; Ricult staff and partner staff collect data using mobile devices; farmers can also provide data via phone app
Data capture methodologies (secondary data)	Integration of new data sets from other CGIAR members; satellite, climate data, weather, pests, etc.	Integration of financial services data, weather and pest data, soil and coffee 'cupping' data	Integration of satellite, weather, use of APIs to access other relevant data; blockchain, credit data

	CIAT/CGIAR	Grameen Foundation	Ricult
Data storage	Cloud-based and CGIAR servers	Cloud-based/ TaroWorks (salesforce)	Blockchain, cloud-based storage
Data analytics	All types of data analytics (descriptive to cognitive)	Descriptive and diagnostic, with more recent use of predictive for credit scoring	Cognitive analytics for day-to-day management
Data sharing	Open data on CGIAR platform; given open sharing policy, data is not monetized. Support services to help others with data analytics or to support migration of others' data to open data is under consideration	Data is private among Grameen and its partners; once partners "own" the data post-project, usually stays private; no monetization of data assets/no open data	Data is proprietary but Ricult sells data to inform other actors in the space (e.g., fertilizer companies)
Data usage	To inform projects and partners with data needed to support farmers; can be through use of tools farmers manage to tools used by partners. To contribute to learning products available for the sector	To inform projects and partners with data needed to support farmers; can be through use of tools farmers manage to tools used by partners; can be both real-time alerts to farmers (for climate events) or when field agents meet with farmers	To inform real-time credit assessments, satellite/weather alerts to farmers, coordinating communication between market players and farmers
Opportunities	• Development of an open data asset management system prototype • Increased use of big data analytics with integration of multiple data sets • New knowledge products that marry plant and human science	• Improve management of years of agriculture data across contexts • Increase use of data analytics to harvest more insights from data	• Business model that responds to farmers' comprehensive needs; provides real-time support with data collected from and on farmer • Flexible farmer profile "framework" approach that can adapt to new contexts • Demonstration of integration of multiple new data technologies
Challenges	• Client protection, privacy considerations • Development of processes and procedures to assist in open data management	• Data ownership after projects have ended can be ambiguous and requires upfront clarification • Partner capacity and interests determine approaches • Funding for innovation can be limited • Sustainability (time and funding	• Financial sustainability (Ricult's hybrid business model is promising) • Still in startup mode of proving model • Scalability is to be determined
Implications/Key considerations for others considering similar work	• Is there a role for open data? • How can data assets be best leveraged past the primary/research stage? • How could participating/drawing on communities of practice regarding standardization of processes/ frameworks to facilitate open data encourage learning and engagement?	• Is there a role for open data? Where would it go/be managed? • Is there a role for monetizing data assets? • Do adequate data science skills exist? Are new data science skills needed? • How do we support continued learning past project stage?	• Can data assets be monetized (if not already)? • What other revenue-sharing models can be considered (as they relate to data assets)? • What regulatory environments influence decisions as they relate to customer privacy?

RICULT CASE STUDY

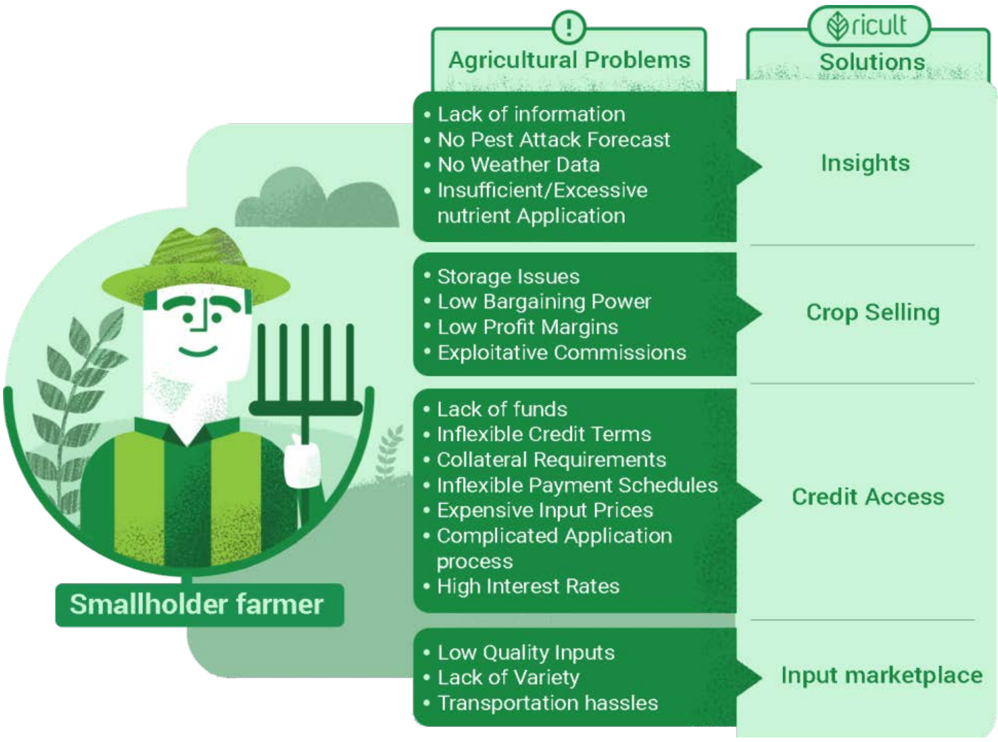
About Ricult

Ricult is a US-based, privately owned farmer-support firm that currently has operations in Pakistan and Thailand. Ricult, whose name comes from agRICULTure, was launched in February 2016. The members of its founding team came from the Massachusetts Institute of Technology and include individuals with diverse backgrounds in agriculture, credit risk modeling, data science, and machine learning and business processes. Ricult aims to strengthen the agricultural ecosystem for smallholder farmers and is focusing on those working in corn and rice. Ricult is considered both a financial technology (fintech) or agritech company and an ecosystem orchestrator.

About Ricult's Model

Ricult aims to support smallholder farmers in a comprehensive way, providing a one-stop marketplace for agricultural support. Figure 9 below, which is drawn from Ricult's website, synthesizes the challenges faced by farmers and how Ricult broadly addresses them.

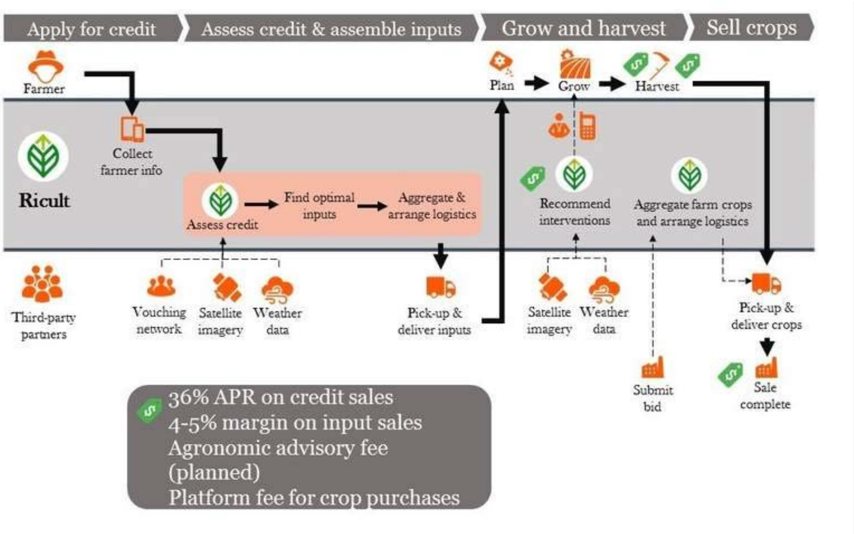
Figure 9: Smallholder farmer challenges and Ricult's solutions



Source: Ricult.com

Ricult utilizes a technology-based integrated solution to address multiple pain points of smallholder farmers. Lack of financing, poor quality and overpriced inputs, exploitation by local money lenders and buyers, and poor agricultural practices (particularly regarding fertilizer and pesticide use), all combine to perpetuate the low yields and low incomes of smallholder farmers. Figure 10 on the following page, synthesizes Ricult's support to the farmer—from initial involvement to the sale of goods.

Figure 10: Ricult farmer experience



Source: https://challenges.openideo.com/challenge/bridgebuilder/ideas/ricult

In Table 5 below, key aspects of the farmer's experience are detailed in parallel with how data is collected from and on the farmer and how it is analyzed and used to support the farmer.

Table 5: Ricult model

	Farmer perspective	Ricult's perspective
Farmer acquisition	Farmers who have less than ten acres of land each are encouraged to form groups of five to ten farmers to access Ricult services. Those with larger landholdings can contact Ricult individually.	Currently, because this is a startup, there is enough demand that Ricult is slowly reaching farmers based on geographic targeting. (They are focusing on areas with fertile soil and farmers who are open to more modern farming techniques, to start the demonstration effect and influence other farmers).
Enrollment/credit application	Farmer(s) meet with a Ricult field agent and start the process by providing their identification information (which is matched with information in national databases). For those who want credit, a field agent assesses willingness to pay via an android tablet with a psychometric assessment. If a farmer receives an adequate willingness to pay score, he or she will be further evaluated with an algorithm that integrates reputational profiling and ability to pay.	To determine credit approval, Ricult combines multiple data sources using cognitive analytics to predict both willingness and ability to pay. A willingness to pay assessment is completed first. If the farmer passes, he/she is further evaluated, which includes reputational profiling and ability to pay. Figure 11 below depicts the components of the credit approval process.

	Farmer perspective	Ricult's perspective
Credit approval/product demand list created	Once credit is approved (or Ricult knows the farmer's plan to pay via cash for inputs), a farmer receives a product demand list that prescribes a set of inputs that will be needed by the farmer to improve his/her yield. This list usually consists of inputs such as fertilizer, pesticides, and seeds. The cost of these inputs dictates the amount of credit they receive (if applicable). Ricult does not give a loan directly to a farmer, but provides the inputs to the farmer, who repays Ricult for the cost of the products.	Some data sources used for the credit approval process (satellite, agronomic data, farmer profile data) are used to determine the product demand lists for both farmers seeking credit and farmers who pay Ricult directly for the inputs. Ricult is not working with any local financial institution in Pakistan but instead has a strategic partnership with a blockchain-enabled distributed ledger company that is creating an international agriculture fund for Ricult.
Inputs	As stated above, farmers can receive inputs directly from Ricult and the inputs are specific to his/her need. Ricult delivers the inputs directly to the farmer's door within a day.	Ricult purchases inputs in bulk from local and international providers at negotiated terms that make it possible to provide higher value goods at a lower cost to farmers. Input suppliers include: Engro fertilizers, Bayer, Guard Rice, DuPont Pioneer, FFC, Maxim and ICI. Ricult uses prescriptive analytics to match information on the farmer's plot to a list of inputs that are developed to predict the farmers specific seed, pesticide, or fertilizer needs.
Information/extension support	In addition to their own field agents, Ricult plans to work with government field agents, input supply companies, and others who already have field agents. Ricult's platform enables other firms to supply information services to farmers. Farmers can receive SMS or outbound IVR messages with crop-specific information such as data on crop health, weather updates, and pest alerts.	Ricult uses data analytics to interpret satellite imagery and microclimate weather data, providing farmers with key insights on issues such as potential crop yield, pest attacks, fertilizer recommendations, and optimal harvesting time. These messages are either delivered directly to a farmer's phone or via extension agents.
Market orchestration	When it is time for farmers to sell their goods, Ricult links them directly with buyers, bypassing middlemen who are known to exploit smallholder farmers by keeping them perpetually indebted.	Current buyers include: K&N, Maxim and others.

Farmer Acquisition

In Thailand, 70 to 80 percent of farmers are women. Men generally work in factories and urban centers. In Pakistan, by contrast, most farmers are men, although this can vary by crop. There are more women cotton farmers, for example, and women are also more involved in batting cotton. According to Usman Javaid, CEO of Ricult (who was interviewed for this case study):

"We have rarely encountered women farmers in Pakistan. However, if farming information is more available to women, the monopoly of men's involvement in agriculture due to their knowledge of the right approaches to take will be diminished. Women can also have access to the information and it gives her a greater chance of trying her luck at farming."

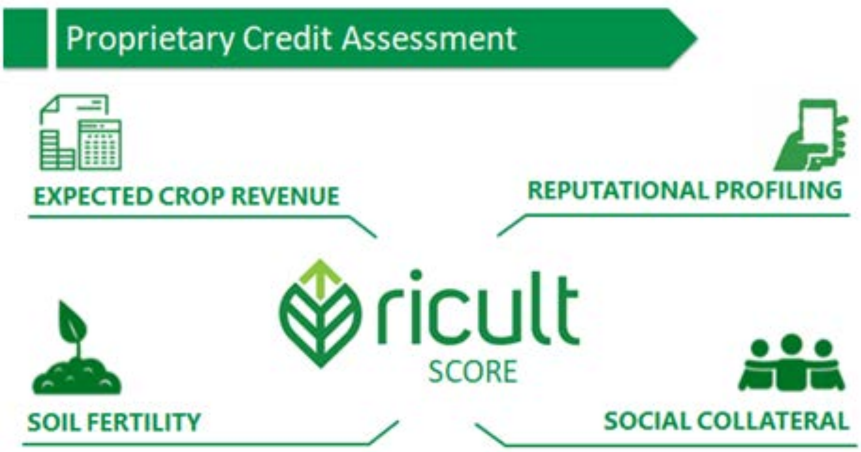
Javaid noted that most of the data Ricult needs to develop farmer profiles can be collected remotely through APIs that connect Ricult with other information databases. Ricult wants to know as much about the farmer and his or her needs as possible, before asking unnecessary questions.

Credit Scoring

Central to the concept of machine learning is the ability to adapt to new information. Ricult's algorithm for a farmer's willingness to pay is based on internally developed psychometric variables (see Figure 11). The algorithm "improves itself" by taking into account prior predictions and outcomes. For example, during the first lending season, Ricult rejected 97 percent of its applicants and experienced moderate defaults. The algorithm used these outcomes as input for the next lending season, resulting in fewer rejections (91 percent) and near zero defaults. In time, this advanced and real-time use of results from more lending cycles will undoubtedly continue to improve the lending process.

Ricult also uses reputational profiling. They invite others in the village who know the borrower to rate him or her on trustworthiness, among other characteristics. "The idea is not to just rate people but to improve their credit worthiness," Javaid said. Currently, Ricult is partnering with a US-based company that created its agricultural fund. Ricult prefers to provide farmers with in-kind support (providing farmers the actual inputs they need) rather than cash loans—which farmers might use for other purposes (and then go to a money lender to get funds for their farms).

Figure 11. Credit scoring process



Source: Ricult, 2018.

Satellite/Weather/Pest Data

Satellite data supports both the credit approval process and the development of a farmer's product demand list. The graphics below (Figure 12) demonstrate how Ricult's system can provide information on a farmer's crop health (even prior to the application process) and how this data can be used to monitor change over time.

Figure 12. Ricult satellite interface

Source: Smith, 2017.

Transfer of Information and Agriculture Extension Support

Usman Javaid noted farmer interactions take place along a spectrum of information sharing. At one end, farmers rely with confidence on their mobile phones; at the other end, they need strong assistance. Ricult's experience has been that most farmers in low-resource settings need to be guided in using new technology until they become comfortable. According to Javaid:

"There is a trust deficit because of the years of scams and exploitation. Any element that is not indigenous to where a farmer comes from takes time and proof that something new will work. Trust will eventually happen, it is a slow process. Farmers need to believe in the technology themselves. You need to initially invest in a few who come to realize the suggestions are reliable. There is a strong word of mouth in the community. A critical mass can start. You convince a few, others will follow."

Javaid noted that the level of smartphone penetration also makes a difference. In Thailand, many more farmers have smartphones and can get data directly on their phones. In Pakistan, smartphone use is less common, and farmers need more assistance using digital technology.

Scaling the Ricult Model

When Javaid and his team developed Ricult, they wondered why they did not see more mega-companies doing similar work. They have since realized that a lot of hard grunt work is required to reach a critical mass. To a certain extent, agricultural extension support needs human intervention (government workers or supply companies) to build trust. But the key is scale. There are low profit margins but immense scale opportunities. The question is how to scale technology to the maximum number of people. Javaid has found that rather than focus on one part of the value chain, the entire value chain must be considered. According to him:

"Many companies are too narrow in their approach and that is where they go wrong. Because the agriculture ecosystem is so integrated, you cannot just focus on one part and assume all the others will follow. Once a farmer is convinced, you do not have to go back to them as often in person as they rely more and trust the recommendations coming through technology."

Ricult believes a scalable model must address the needs farmers have in common but must also be flexible. For example, Ricult uses specific parameters to assess people's ability and willingness to pay. They start from a customized set of psychological questions, but honesty and willingness to pay metrics may vary from country to country. The data sources Ricult uses can also vary depending on the country. In Pakistan, they rely on certain national databases, whereas in Thailand, other data sources may be available. To reach their objectives, flexibility is key.

Sustainability

Ricult is a for-profit company that has multiple revenue streams: interest paid on credit, commissions on farmer inputs, fees paid by crop buyers who use Ricult's platform, monetization of their data by other local actors, and investment funds that provide start-up capital for developing and fine-tuning their model.

The Future

Ricult envisions additional innovation ahead. In an interview by Techsauce.com, Ricult co-founder and Chief Product Officer Aukrit Unahalekhaka noted their current use of machine learning and artificial intelligence to analyze data for farmers may soon be augmented by drones. Looking ahead, they plan to integrate the use of drones so that a farmer can open the Ricult app, press a button, and receive drone measurements of farmland and weather conditions. They are also considering the use of virtual reality.

Javaid sees more changes ahead:

"I see a future where you will have digital currency driving rural economies. I believe that there will be a cashless economy that will be fueled by blockchain and digital currency. We are already working on this... It is going to come whether people like it or not. The issue is that most microfinance institutions and banks in fintech face challenges of scale. The future will be based on data-driven decisions and digital currency built on blockchain."

However, Javaid noted that a government's ability to encourage private actors to build blockchain solutions will also affect the pace of this transition. Javaid continued:

"When people become aware that data can be monetized, they will make it harder to get data. There will be a time when people will charge companies if they want the farmer's personal data. I am a strong supporter of this. I think people who provide data should be compensated. [The farmer] drives the data that is provided, he monetizes and uses it for his own benefit. I see an empowered farmer in the future."

Ricult Case Study References

Dickie, S. (2017). Agritech: Ricult's Co-Founder explains what this innovative start up plans to do for the Thai farming industry." Retrieved from: https://techsauce.co/en/asia/agritech-ricults-co-founder-explains-innovative-start-plans-thai-farming-industry

Interview with Usman Javaid, CEO of Ricult. January 2018.

Ricult website: http://www.ricult.com

Video on Ricult–Pakistan: https://www.youtube.com/watch?v=DwHNHyYPdRg

Video on Ricult–Thailand: https://www.youtube.com/watch?v=a3ERjkAAHsE

GRAMEEN FOUNDATION CASE STUDY

About Grameen Foundation

Grameen Foundation USA (Grameen) is a global nonprofit organization with a mission to enable the poor, especially women, to create a world without hunger and poverty. Grameen works at the intersection of agriculture, financial services, and health, using technology and partnerships to create breakthrough solutions for smallholder farmers and poor communities in Africa, Asia, and Latin America.

Grameen and Digital Technology Solutions

In the past ten years, Grameen has evolved as a technical service provider and leader in digital innovation for smallholder farmers. Its innovations have paralleled the evolution of digital technology for agriculture more generally. With each software iteration that Grameen designs and pilots, it incorporates lessons learned and integrates its applications with emerging technology for bundled farm-to-market services and smarter data usage. Its latest digital innovations incorporate satellite and geodata. Common to many of Grameen's solutions is the use of TaroWorks, a mobile- and cloud-based technology originally developed by Grameen for data storage and analytics.

Table 6 below provides a summary of select Grameen programs, the technologies used to capture data, and different techniques in analysis and application of that data for farmer benefit.

Table 6: Progression of farmer profiling in Grameen programs

Name of Project	Country	Technology	Data usage
Community Knowledge Worker (CKW) Program	Uganda	Custom development: mobile- and cloud-based technology for data storage and analytics	Collected farm-level data for profiling customers and assistance provided with GAP, including identifying and treating production issues such as pests and diseases
Farm Development Plan with UTZ/ Rainforest Alliance and Mars	Indonesia	Android- and cloud-based technology; diagnostic analysis and rules-based predictive tools that are applied to assist a farmer's decision making	Advisory services for productivity increase, production quality, farm investment, and certification using individualized farm development plan with monthly activities and costs and annual projected profit and loss
Agrotech	Ghana	Mobile- and cloud-based technology for data storage, SMS, radio, and video tools incorporated for farmer communication and education	Provided GAP recommendations and individualized farm management plans Developed algorithm that used farmer data to cluster farmers into four groups based on experience; a "Know Your Client" tool for the field agent
FarmerLink	Philippines	Combines satellite data and farm-level observations collected using TaroWorks mobile technology Technology service providers included aWhere for weather alerts and engageSpark for SMS/IVR messages	Assessment and monitoring of GAP adoption. Common platform for multiple value chain players (government support agency, commercial purchaser, financial services provider) to interact with farmer data Early Warning System that provides alerts directly to the farmer's phone for extreme weather events and potential pest and disease outbreaks using farmer's GPS location; SMS alerts sent to farmers' phones that contain the warning and a practical recommendation that can be adopted in the farm Automated diagnostic analysis incorporated to assess certification eligibility and corrective action recommendations

Name of Project	Country	Technology	Data usage
mSourcing Solution	Colombia	Salesforce app, cloud- and farm-level data collected using TaroWorks mobile technology	Increased quality and connection of farmers' associations to markets by using data on inputs provision, farming activities, and crop cycle to forecast and provide real-time information on farmer supply capacity; data allowed better integration with businesses such as grocery store; integrated back-end functions and business intelligence tools, allowing buyers to better manage orders, inventory, payments, and client billing
Agricultural Risk Evaluation Tool (ARET)	Colombia	Salesforce app, cloud- and farm-level data collected using TaroWorks mobile technology SMS data capture as well as CKW search application that enabled CKWs to quickly answer farmers' questions and track searches for crop information to glean further insights	Connection to financing and detailed financial plan; predictive analysis to determine farmer's propensity to pay; integrated view of farmer's contact information, farming practices and productivity, credit history and score (when available), inputs/purchases, farm investment needs, and improvement plan targets
Soil sampling			Precision soil sampling, data collection, and data analysis for recommendations leading to better utilization of fertilizers and other soil amendments, determining the economic threshold for treating pest and weed infestations
SAT4Farming	Ghana	Satellite data integrated with mobile application for GAP recommendations and development of individualized farmer development plans	Improve land and water usage; incorporate satellite data to increase reach and improve accuracy and effectiveness of diagnostic and monitoring activities

Note: for more information on the projects listed in the table, please visit https://www.grameenfoundation.org/resources/publications

Data Capture

Grameen collaborates with private, public sector, and NGO partners to leverage technology and community-based networks to promote two-way information flows that allow farmers to make informed decisions, facilitate financial transactions, and create more inclusive value chains. A typical farmer profile across the different projects includes the domains of information found in Table 7 on the following page.

Table 7: Type of data collected

Type of information	Data collected
Know your customer	Name, gender, age, marital status, geographic location (both physical address information and GPS locations of the farm), relationship to household head, number of household members, mobile phone telephone number
Poverty status (Poverty Probability Index [PPI][3])	Depending on country, a current PPI survey is used
Land ownership	Including hectares and farm description; mapped farm data that plots size of land holding (use of GPS to map landholding), access to water, proximity to protected land
Financial services	Whether household has bank accounts, has received credit in last 12 months; some projects track balances, M-Pesa (mobile money), and bank inflows and outflows for the last six months, layaway metrics
Crop information	Amount of land under production, primary cash crop, secondary crops, soil conditions
Inputs used	Purchase of inputs, use of specific agro-input dealers
Capture of information on GAPs	Depending on the crop, GAPS are monitored over time to track their partial or full adoption by farmers (e.g., adoption of a specific fertilizer or pesticide practice, mulching)
Capture of surveillance	Plots have been mapped with GPS, but weather and pest data are collected with support of satellite/weather provider
Capture of marketing information	Market prices, market days, traders, harvest production and productivity for sustainable sourcing, adherence to certification, and/or market standards
Supply chain management	Expected harvest amounts, field activities schedule, inputs provision, seedlings losses, product losses, deviation between production forecasted and harvested, quality of production, certifications achieved

Extending Data Use

Grameen Foundation began the basics of digital on-farm data collection nearly a decade ago to improve efficiency of extension services. Over the years, it has sought ways not only to improve on-farm production but also to leverage the information collected and the resulting technology insights to improve the farmer's access to finance and market, while incorporating new tools and analysis techniques.

Improving access to finance

Grameen Foundation joined forces with *Cooperativa de los Andes* (delosAndes), a large coffee cooperative in Antioquia, Colombia, to pilot a new way to collect data on their coffee farmers, to provide extension services by leveraging a CKW network, and to increase the farmers' borrowing potential. When Grameen and delosAndes first partnered, delosAndes was using paper-based surveys and entering their data into Excel spreadsheets. Grameen assisted delosAndes to migrate their farmer profile data to TaroWorks.

With the data collected from the initial farmer profile and ongoing monitoring, farmer GAPs were monitored to track adoption rates and yields. Grameen and delosAndes developed a shared dashboard that tracked this information. To create additional value for this data, Grameen led the creation of a tool to help assess a farmer's propensity to honor her loan commitments, using alternative data sources to build a credit score. The tool was named Agricultural Risk Evaluation Tool (ARET). With ARET, Grameen aggregates and merges data sources available to its partner, uses chi-square, correlation, and t-tests to identify variables associated with loan repayment, and uses the classification and regression trees (CRT) algorithm to build the final version of the risk model. The final prototype draws on fewer than ten variables (primarily related to on-farm behaviors) and classifies farmers across a risk profile. This project is still underway, with further adjustments to the variables in the algorithm. Ongoing fine tuning of alternative credit scoring or risk methodologies by financial and nonfinancial firms is an expected process. See Figure 13 for the shared dashboard of ARET.

3 In 2017, the name of the Progress out of Poverty Index (PPI) was changed to the Poverty Probability Index. For further information about the index see: https://www.povertyindex.org/about-ppi.

Figure 13: ARET shared dashboard

Source: Grameen Foundation

Grameen also worked with Musoni Kenya to develop an agricultural financing product known as Kilimo Booster to provide credit for agricultural activities and provide farmers with basic agricultural support information. Together, Grameen and Musoni built a tablet-based, digital field application for customer registration and cash-flow analysis. With digital data collection, Musoni improved the accuracy of the cash-flow estimate and shortened the time required for the loan application process. Musoni builds the profile of the farmer over time; the loan officer-facilitated interviews collect updated data during each new loan application.

A client-facing application was also developed that allows farmers to query the Musoni system directly for their loan balances and records of their most recent statements, so that they don't need to rely on information from their loan officers.

Improving quality and access to markets

Farmers can gain additional insights from the data collected to maximize their selling positions. Information might include certification eligibility, quality assessments, and direct connections to market.

Grameen Foundation worked with Coocafisa Cooperative in Colombia to integrate the use of technology and the CKW model to train farmers using tablets with the agronomy tools from C.A.F.E. Practices, provided by the Farmer Support Center in Colombia. In this project, "cupping" data was integrated into the farmer profile data. The data (a quality level assigned to a farmer's coffee beans by a coffee tasting lab) was useful for both international and national marketing purposes. A technician picked up the coffee samples and sent them to a lab; the coffee tasting lab assigned a quality designation (aligned with coffee standards of the Specialty Coffee Association of America, for example), and input this data into the system. This assisted the technician in identifying where quality of the bean may have suffered (wet-mill processing? during transport?), considering other data on the farmer and his or her adoption of GAPs for coffee. The farmer and the cooperative took corrective action.

A more direct example of connecting buyers and farmers is Grameen Foundation's M-Sourcing tool. As part of its partnership with Salva Terra, a nonprofit that promotes organic farming in vulnerable communities in Antioquia, Colombia, Grameen leveraged a shared farmer profile platform so that Salva Terra technicians could collect and disseminate information to help farmers improve the quality of their products, increase their outputs, and connect to large-scale buyers.

The m-sourcing platform uses farmer profile data to forecast and provide real-time information on a farmer's supply capacity to businesses such as grocery stores. The back-end functionalities allowed Salva Terra to manage inputs supply, orders and deliveries, inventory, payment, and client billing.

Leveraging analysis techniques to provide specific, timely and actionable information

Grameen Foundation seeks to incorporate both leading edge technology and analysis techniques to improve farmer interaction. An example of this is the use of simple SMS technology to create timely alerts.

FarmerLink is a multi-sectoral collaboration of Grameen Foundation and its implementing partners, the Philippine Coconut Authority (PCA, a government agency), Franklin Baker Company of the Philippines (coconut buyer), and People's Bank of Caraga (financial services provider). Field agents are all equipped with mobile agricultural extension tools to help monitor farmer progress, promote GAPs, and deliver financial advice. The program also includes technology service providers—aWhere and engageSpark—who developed an Early Warning System. Grameen Foundation worked with researchers to identify the conditions leading to pest and disease outbreaks and analyzed big data forecast information to develop advisories based on the GPS location of the farms. A feedback mechanism to collect on-farm observations corresponding with the alerts will be used to develop machine learning algorithms, providing continuous system improvement.

In Indonesia, similar to FarmerLink, Grameen Foundation and UTZ/Rainforest Alliance have worked with Mars Incorporated to apply years of scientific research acquired by Mars to develop action plans for farmers. Farmers are expected to increase their production threefold as a result of adhering to good agricultural practices. Large amounts of historical data were synthesized into diagnostic algorithms that lead to specific recommendations for monthly activities and costs (projecting farmer household income for decades—corresponding to the life of cocoa trees). The resulting plans are extremely detailed and personalized, based on farm size, conditions, and the farmer's agricultural practices. A new iteration of this Farm Development Plan will incorporate satellite data in Ghana for diagnostic and monitoring activities to increase precision and reach.

Lessons Learned

Across these different projects, Grameen has had the opportunity to advance different aspects of the farmer ecosystem. Efforts have provided insights into the challenges of engaging organizations having little technological experience in roles within a new digital ecosystem. A few of these lessons are highlighted below.

Sustainability

Grameen Foundation has been engaged in digital innovation for agriculture since 2009 when ICT for Agriculture was still nascent. Its first program was the **CKW program** in Uganda that used a unique extension model to address many of the challenges and weaknesses of other extension and advisory services (EAS) programs. Challenges include low illiteracy; limited resources for government EAS provision; poor communication among farmers, advisors, and researchers; and lack of hands-on assistance and interpretation of agricultural information sent via ICTs. Grameen Foundation developed iterations of CKW in other countries, incorporating valuable lessons along the way.

One of the most challenging obstacles to project-based extension initiatives is sustainability, especially after initial grant funds have been exhausted. Grameen Foundation planned the CKW program with sustainability in mind; however, after the funding ended, Grameen lost its public sector partner and could not find a private sector partner to own and maintain the program.

With the ongoing iterations of the CKW program, Grameen Foundation has expanded the definition of the CKW to include agents of both public and private entities, volunteer or professional, paid and unpaid. To ensure sustainability of the platform, for example, Grameen Foundation inserted a change management approach into the project with delosAndes to examine their technology, processes, and people. This entailed examining front- and back-end technology infrastructure, marketing strategies, and the roles and responsibilities of staff intersecting with the product. In Grameen's FarmerLink program, the Philippine Coconut Authority (a government entity) is a key partner. It owns the platform for collecting data on farmers and directs critical information and support to them through field agents and SMS/IVR messaging based on weather, pest, and satellite data.

Digitization

Depending on the partner, the process of training extension agents and staff members requires extensive hand-holding and refreshers as the technology is implemented or updated. Most staff in typical service provider organizations have very basic understanding of technology and are often averse at first to incorporating mobile digital technology tools into their work. However, over time, field-level staff have noted the reduced workload and the immediate guidance it gives them in providing specific support to a farmer. Some partners have also made the data collection process part of their staff appraisal/review process, to ensure quality of data.

CKWs, who are themselves typically farmers in local program communities, are initially motivated by gaining access to and capacities in new technologies, such as smartphones, which aid them in communication and knowledge sharing. After being trained to use a project's tools—including interactive guides and data collection and content library tools—CKWs see themselves as valuable leaders and extension agents, providing tangible benefits to their neighbors and acquiring increased recognition in their community.

Gender

Most of the projects noted in this case study were implemented within the context of a partner's operations. Most initial data analyses were disaggregated by sex (male vs. female) and in some cases, compared male-headed households to female-headed households in terms of GAP adoption or differences in uptake or outcomes. Grameen is currently revising how it surveys farmers to better understand the role that a person may play on the farm. For example, a woman might be interviewed about the farm because her husband is not at home at the time of the interview, but she might or might not be the farmer, decision maker, or plot manager.

Looking Forward

Grameen Foundation's ownership of data in these programs may be short lived. This prompts questions about the role of open data. Given the in-depth profiles of farmers collected, it would be unfortunate if the data did not create value beyond a specific project. Going forward, Grameen is seeking to find mechanisms that will make its data more findable, accessible, interoperable, and reusable.

Grameen Foundation Case Study References

Food and Agriculture Organization. (2017). Grameen CKW data visualization with Palantir. Retrieved from: https://www.youtube.com/watch?v=CKHLXAMFSQU

Gatti, G., & Hipolito, T. (n.d.) Bringing coconut farmers into the 21st century through mobile agriculture. Retrieved from: https://foodtank.com/news/2017/02/farmerlink-mobile-agriculture-philippines/

Grameen Foundation website on agriculture: https://www.grameenfoundation.org/what-we-do/agriculture

Lessons Learned, 2009-2014: Community knowledge worker Uganda program. (October 2014). Retrieved from: https://www.grameenfoundation.org/resource/lessons-learned-2009-2014-community-knowledge-worker-uganda-program

CGIAR'S PLATFORM FOR BIG DATA IN AGRICULTURE

About CIAT and CGIAR

CGIAR, formerly known as the Consultative Group on International Agricultural Research, is a global research partnership whose goal is to reduce poverty, improve food and nutrition security, and improve natural resource and ecosystem services. CGIAR is a consortium of 15 research centers that are independent entities with specific mandates including research on tropical, semi-tropical, and semi-arid agriculture; wheat, maize, rice, and potatoes; and food policy.

About the CGIAR Big Data Platform

CIAT, or the International Center for Tropical Agriculture (CIAT for its Spanish acronym), and IFPRI, two of the 15 CGIAR research centers, launched the CGIAR Big Data Platform in late 2017. CGIAR identified five data trends they believe should be harnessed and provide impetus for the development of the Platform:

1. growing use of mobile phones to both collect and send data
2. improvements in remote-sensing data availability and resolution
3. rapid spread of cheap sensor technologies that enable the Internet of Things business models and new approaches to on-farm measurement
4. distributed computing and storage that democratizes analytical power
5. many innovations popping up that build ecosystems that support start-up businesses and spur the development of new digital model

> "The CGIAR Platform for Big Data in Agriculture is where information becomes power: power to predict, prescribe, and produce more food, more sustainably. It democratizes decades of agricultural data empowering analysts, statisticians, programmers, and more to mine information for trends and quirks and develop rapid, accurate, and compelling recommendations for farmers, researchers, and policymakers." – CGIAR Platform for Big Data in Agriculture; https://www.cgiar.org/research/program-platform/platform-for-big-data-in-agriculture/

The Platform embraces big data, which CGIAR defines as "open, harmonized, interoperable, and integrated datasets from multiple domains aimed to accelerate agricultural research and data use in service of development goals." CGIAR emphasizes that it does not define big data by the size of the data, but whether it generates new insights that were previously impossible. CIAT has conducted data analytics with as few as 70 records coming from field trials and 50,000 records on farmers. **The capacity to analyze information is making the difference.**

The aim of the Platform is to increase the impact of agricultural development by embracing big data approaches—opening and sharing agricultural data, creating data assets that are global public goods, developing new methodologies, and brokering big data innovation.

To operationalize its own definition of big data, CGIAR is:

- **Making its data open:** CGIAR is a member of the Global Open Data for Agriculture and Nutrition (GODAN) partnership. In 2013, all its research centers signed on to GODAN's Open Access and Data Management Policy. As a result, the Big Data Platform is investing in all centers' progress towards more open research—making it more Findable, Accessible, Interoperable, and Reusable (FAIR).

- **Ensuring its data is harmonized and interoperable:** Five communities of practice (CoPs) within the Big Data Platform seek to model effective use of harmonized and interoperable data technical disciplines. The five CoPs have the following missions:
  - *Crop modeling:* This CoP aims to improve the global coordination of crop modeling efforts in agricultural research among CGIAR centers and collaborators and create new knowledge about suitability of models to the challenges they address.
  - *Data-driven agronomy:* This CoP develops big data tools based on three main principles: 1) increased use of observational information, 2) data mining, and 3) contextualized information. It brings CGIAR's strategic partners and CGIAR centers together to collectively develop and implement big data tools based on FAIR principles.
  - *Geo-spatial data:* This CoP undertakes activities to bring spatial scientists together through coordinated communication, community-developed products, and convening members at various events to represent CGIAR in this area of expertise.
  - *Ontology:* This CoP establishes best practices and guidelines in the selection, use, and application of semantics for data harmonization at the levels of collection and storage, for data interoperability and data discovery following the FAIR principles.
  - *Socio-economic data:* This CoP brings together actors interested, able, and willing to tackle major issues related to socio-economic (survey) data and moving forward to develop ontologies and make the data interoperable.
- **Developing a platform to leverage integrated data sets:** To make CGIAR consortium data available, the Big Data Platform includes a prototype data harvester that enables searches across all CGIAR databases. Users can conduct simple keyword searches for information from relevant datasets and publications. The harvester API allows the user to pull data for their own research. The vision is that researchers will more easily discover and assemble commonly-used data for food security (e.g., the World Bank Living Standards Measurement Survey) alongside other CGIAR and non-CGIAR data sources to unlock new uses and analyses. Brian King, the Coordinator for the Big Data in Agriculture Platform, shared, "We want to demonstrate new leading edge uses of data, leveraging the data we have to better interpret and use a vast and growing constellation of digital data to understand and respond to the world in real time." Brian believes that the CGIAR network is uniquely positioned to provide a solid set of tools and resources to accelerate digital transformation in the agriculture sector.

Challenges: Privacy, Interoperability, Sustainability

A future goal of the Platform is data interoperability and meta-analysis at a global scale. Discrete successes show how this can be done (see the Aclimate Colombia example below). Interoperability of data must be approached responsibly. Given the importance of data privacy, the Big Data Platform is researching encrypted data-storage and -sharing tools that can enable responsible use of sensitive data. For example, the use of a specific farmer's GPS location may be protected, while aggregate data including this information could be shared. While the debates on privacy and data sharing continue, practical steps will require learning more about what true informed consent means when working with digital data. The Platform for Big Data representatives are drafting responsible data guidelines that aim to balance ethical concerns with the desire to unlock data sharing in the agriculture sector.

Another challenge related to interoperability is the need to standardize the way data is shared. Daniel Jimenez, the Platform leader of the Data-Driven Agronomy CoP, was interviewed for this case study. He shared that his team "…spends 80 percent of its time in data curation and cleaning. To match and integrate separate data sources is complicated, so the ontologies we have developed over the years and the ones we continue to build are important but they are also a work in progress." They are aiming for an agile, function-driven and iterative approach to sharing and using data with the goal of integrating data from multiple sources— be they farm trials, breeding programs, predictive or geospatial analytics, and farmer profile data collected by local support organizations. The goal is to provide farmers with individualized, timely, and concise support.

Finally, sustainability is a continual challenge. There has been a growing recognition that private sector involvement is critical. While decades of agriculture research have been conducted, CIAT, like many of the other CGIAR research centers, is a nonprofit working to create open public goods. Once a project ends, researchers may not be able to continue to work with the farmers who participated in the original research. But this is where the private sector can provide a critical path to scale, impact, and the sustainability of the benefits of the work started by CGIAR researchers.

Big Data Analytics is a Paradigm Shift

According to Jiminez:

"What is most exciting about the move toward big data analytics is that data-driven analysis is a paradigm shift for technicians in the field as well as for farmers. In the past, a technician might use on-farm information to detect patterns and you develop a recipe for growing maize, for example. If you are lucky, you use data to confirm what you are learning, but then one day you discover there are three other factors that were actually influencing yields, like amount of available sunlight during specific moments of the plant's development. Data-driven analytics is causing us to think about factors that we have not thought about before."

Jimenez believes that satellite and remote-sensing data, along with climate information, will eventually satisfy the need for much of the data required from the farmer.

But Jimenez cautions that while automation of data collection, use, and dissemination are exciting, the interpretation of results cannot be automated. Machine learning models generate information based on the data you input into the model. He explained:

"If you do not incorporate human insights into the model, the data can give you the wrong conclusion. Take, for example, one project in Nicaragua. We found through the data that the increase in nitrogen applied to a plant reduced production. This didn't match our expectations. You learn from those on the ground that when a plant gets more nitrogen, it is more attractive to pests. However, 'pests' were not incorporated into the original model, so the conclusion was that a farmer should reduce nitrogen application, which makes no sense."

In addition to the launch of the Platform, CGIAR is aiming to inspire the sector with innovations by supporting pilot projects that make use of existing CGIAR data to develop novel approaches that use data-driven insights to inform policies and help farmers lead better lives. Through their CGIAR Big Data in Agriculture Inspire Challenge, they recently invested in five teams that will be innovating around the four Inspire Challenge categories: Revealing Food Systems, Monitoring Pests and Diseases, Disrupting Impact Assessment, and Empowering Data-Driven Farming.

Aclimate Colombia: A mini-case

The Aclimate Colombia project was a cross-sectoral partnership led by CIAT that engaged the Colombian Ministry of Agriculture and Rural Development (MARD); the National Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM); and Fedearroz, an association of rice growers. Fedearroz shared with CIAT the data they had gathered over a 25-year-period conducting rice surveys, monitoring harvest results, and collecting plant science data. IDEAM shared station-level weather data germane to rice growth (e.g., temperature, precipitation, humidity, and solar radiation). The Aclimate platform aggregates and packages these datasets for use in various data analytics.

The Aclimate initiative very quickly produced analytics that predicted a major dry period would disrupt the upcoming growing season. The appropriate response was to delay planting. Equipped with this information, Fedearroz designed and broadcast communications for rice farmers about the ideal time frames for planting. The analytics also revealed important regional distinctions in rice productivity; some regions were mostly dependent on average minimum temperature while others were mostly dependent on amount of solar radiation or frequency of rainfall. Many farmers used these important and regionally specific insights to guide their decision making, leading to yields that were significantly higher than those of farmers who ignored the Aclimate recommendations.

This case study shows how new use of existing data can provide granular insights and spur progress in the field. A study conducted by the Open Data Institute to assess the impact of this open data platform estimated that improvements in decision making by farmers armed with the new insights led to prevention of potential crop losses valued at $3.6 million for the agricultural sector as well as the Colombian economy.

CGIAR Case Study References

Aclimate Colombia. http://www.aclimatecolombia.org/

Africa: CGIAR wants to build a big data platform and innovate with start-ups to help smallholder farmers. (2017). Retrieved from: http://allafrica.com/stories/201710090620.html

CGIAR. Event Summary and Agenda for Action, First Annual CGIAR Convention on Big Data in Agriculture: Alliance for a Data Revolution, September 19-22, 2017. Retrieved from: http://bigdata.cgiar.org/wp-content/uploads/2017/11/First-Annual-CGIAR-Platform-for-Big-Data-in-Agriculture-Convention-Summary-Report.pdf

CGIAR. CGIAR Open Access and Data Management Policy. (2013). https://cgspace.cgiar.org/bitstream/handle/10947/4488/Open%20Access%20Data%20Management%20Policy.pdf?sequence=1

CGIAR. (2017). Why Big Data? Retrieved from: http://bigdata.cgiar.org/about/

Carolan, L., Smith, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J., & Gerry W. (2015). "How can we improve agriculture, food and nutrition with open data? Global Open Data for Agriculture and Nutrition (GODAN) and Open Data Institute. http://www.godan.info/sites/default/files/old/2015/04/ODI-GODAN-paper-27-05-20152.pdf

Community of Practice on Crop Modeling. (2017). CGIAR Platform for Big Data in Agriculture. CGIAR, CIAT, IFPRI. http://bigdata.cgiar.org/communities-of-practice/crop-modeling/

Community of Practice on Data-Drive Agronomy. (2017). CGIAR Platform for Big Data in Agriculture. CGIAR, CIAT, IFPRI. http://bigdata.cgiar.org/wp-content/uploads/2017/05/2017-Data-Driven-Agronomy-CoP-Workplan.pdf

Community of Practice on GeoSpatial Data. (2017). CGIAR Platform for Big Data in Agriculture. CGIAR, CIAT, IFPRI. http://bigdata.cgiar.org/wp-content/uploads/2017/05/GeoSpatial-Data-Workplan.pdf

Community of Practice on Ontologies.(2017). CGIAR Platform for Big Data in Agriculture. CGIAR, CIAT, IFPRI. http://bigdata.cgiar.org/wp-content/uploads/2017/05/CGIAR-Ontology-CoP_-Workplan-2017.pdf

Community of Practice on Socio-Economic Data. (2017). CGIAR Platform for Big Data in Agriculture. CGIAR, CIAT, IFPRI. http://bigdata.cgiar.org/wp-content/uploads/2017/04/1.-2017-Socio-Economic-Data-CoP-Workplan-1.pdf

International Center for Tropical Agriculture (CIAT). CGIAR. http://www.cgiar.org/about-us/research-centers/international-center-for-tropical-agriculture-ciat/

Interview with Brian King, Coordinator for the Big Data in Agriculture Platform, CIAT. December 2017.

Interview with Daniel Jimenez, Big Data Team Leader for CIAT and facilitator of the Data-Driven Agronomy CoP in the Platform. January 2018.

McDade, M. (2017). Congratulations to our Inspire Challenge finalists. Platform for Big Data in Agriculture. CGIAR. http://bigdata.cgiar.org/congratulations-to-our-inspire-challenge-finalists/

Young, A., & Verhulst, S. (2017). Aclimate Colombia: Open Data to Improve Agricultural Resiliency. Open Data Impact. http://odimpact.org/files/case-aclimate-colombia.pdf

BIBLIOGRAPHY

Accenture delivers innovative agriculture and healthcare solutions via Karnataka MobileOne—India's first mobile governance platform. (2015). Accenture. Retrieved from https://www.accenture.com/t20150723T213622__w__/in-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_10/Accenture-Delivers-Innovative-Agriculture-Healthcare-Solutions-Karnataka-Mobile-One.pdf

Afadhali, J. P. (2016). Agrilift starts crop monitoring tool to boost food production - The East African. Retrieved March 22, 2018, from http://www.theeastafrican.co.ke/Rwanda/Business/Agrilift-starts-crop-monitoring-tool-to-boost-food-production/1433224-3483852-23ky76z/index.html

Africa: CGIAR Wants to Build a Big Data Platform and Innovate with Start-Ups to Help Smallholder Farmers - allAfrica.com. (2017). Retrieved March 23, 2018, from http://allafrica.com/stories/201710090620.html

Agricultural Weather and Agronomic APIs Platform. (2018). Retrieved March 22, 2018, from http://developer.awhere.com/api/reference

Agriculture; plantations; other rural sectors. (2018). Retrieved February 14, 2018, from http://www.ilo.org/global/industries-and-sectors/agriculture-plantations-other-rural-sectors/lang--en/index.htm

Anderson, J. (2018, January). Interview with CGAP.

Anderson, J., Marita, C., & Musiime, D. (2016). National Survey and Segmentation of Smallholder Households in Tanzania Understanding Their Demand for Financial, Agricultural, and Digital Solutions. CGAP. Retrieved from http://www.cgap.org/sites/default/files/Working-Paper-Smallholder-Survey-Tanzania-May-2016.pdf

Andrews, D. (2017). Predictive Model Building 101. Predictive Analytics and Futurism. Retrieved from https://www.soa.org/Library/Newsletters/Predictive-Analytics-and-Futurism/2017/june/2017-predictive-analytics-iss15-andrews.aspx?__taxonomyid=4294967574

Asafu-Adjaye, N.Y., Quaye, W., Yeboah, A., Osei, C., & Agbedanu, E. E. (2017). Appraisal of the Agro-Tech Smart Extension Model in Ghana, Payment options and Challenges in ICT-enabled extension services delivery. *International Journal of Agricultural Education and Extension,* 3(2), 072–084.

Asfaw, S., & Maggio, G. (2016). Gender integration into climate-smart agriculture: Tools for data collection and analysis for policy and research. FAO. Retrieved from http://www.fao.org/3/a-i5299e.pdf

Barrie, G. (2017). How we closed our gender gap in 5 days. Retrieved March 22, 2018, from https://medium.com/farmink/how-we-closed-our-gender-gap-in-5-days-3813fc86e0ca

Barrie, G. (2018). Email exchange with Farm.ink.

Barton, D., & Court, D. (2013). Three keys to building a data-driven strategy. Retrieved March 26, 2018, from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/three-keys-to-building-a-data-driven-strategy

Berne, D. (2018, February). *Data-driven agriculture: An Overview.* Presented at the Webinar hosted by Global Forum on Agriculture Research. Retrieved from https://www.youtube.com/watch?v=bsicKqHZIz4&feature=youtu.be

Bhadra, H. (2017, December). Interview with Accenture.

Big data changing the way businesses compete and operate. (2014). EY. Retrieved from http://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/%24FILE/EY-Insights-on-GRC-Big-data.pdf

Big Data Roadmap: Accelerating Business Returns by Anchoring Your Big Data Journey. (2017). Think Big Analytics. Retrieved from https://www.thinkbiganalytics.com/wp-content/uploads/2017/02/Think-Big-Analytics-Roadmap-Data-Sheet.pdf

Biggs, P., Garrity, J., LaSalle, C., Polomska, A., & Pepper, R. (2016). Harnessing the Internet of Things for Global Development. ITU and Cisco Systems. Retrieved from https://www.itu.int/en/action/broadband/Documents/Harnessing-IoT-Global-Development.pdf

Blockchain: Beyond Bitcoin to Agriculture. (2017). Retrieved March 22, 2018, from https://gro-intelligence.com/

Blockchain for Development: Emerging Opportunities for Mobile, Identity and Aid. (2017). GSM Association. Retrieved from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/12/Blockchain-for-Development.pdf

Boyera, S., Addison, C., & Msengezi, C. (2017). Farmer profiling - Making data work for smallholder farmers. The Technical Centre for Agricultural and Rural Cooperation (CTA). Retrieved from http://www.cta.int/images/Farmer%20profiling%20-%20Making%20data%20work%20for%20smallholder%20farmers.pdf

Breul, P., & Chassin, L. (2018, January). Interview with PHB Development.

Building the ICT Infrastructure of MUIIS for SMS Delivery. (2017). MUIIS. CTA. Retrieved from http://muiis.cta.int/2016/12/21/building-the-ict-infrastructure-of-muiis-for-sms-delivery/

Building a Team for Big Data Success. (2018). Retrieved from http://bigdata.teradata.com/US/Big-Data-Quick-Start/People---Roles/

Burke, M., & Lobell, D. B. (2017). *Satellite-based assessment of yield variation and its determinants in smallholder African systems.* Proceedings of the National Academy of Sciences, 114(9), 2189–2194. https://doi.org/10.1073/pnas.1616919114

Byrum, J. (2017). The Challenges for Artificial Intelligence in Agriculture - AgFunderNews. Retrieved March 23, 2018, from https://agfundernews.com/the-challenges-for-artificial-intelligence-in-agriculture.html

Caire, D., Camiciotti, L., Heitmann, S., Lonie, S., Racca, C., Ramji, M., & Xu, Q. (2017). Data Analytics and Digital Financial Services. Mastercard Foundation and IFC. Retrieved from https://www.ifc.org/wps/wcm/connect/22ca3a7a-4ee6-444a-858e-374d88354d97/IFC+Data+Analytics+and+Digital+Financial+Services+Handbook.pdf?MOD=AJPERES

Camp, H. (2016). Getting Early Warning Weather Alerts Where They're Needed. Retrieved March 23, 2018, from http://blog.awhere.com/getting-early-warning-weather-alerts-where-theyre-needed

CGIAR. (2017a). Event Summary and Agenda for Action, First Annual CGIAR Convention on Big Data in Agriculture: Alliance for a Data Revolution, September 19-22, 2017. Retrieved from http://bigdata.cgiar.org/wp-content/uploads/2017/11/First-Annual-CGIAR-Platform-for-Big-Data-in-Agriculture-Convention-Summary-Report.pdf

CGIAR. (2017b). CGIAR Platform for Big Data in Agriculture: Inspire Challenge Proposal. Retrieved March 22, 2018, from http://bigdata.cgiar.org/wp-content/uploads/2017/09/Diksha-Arora-cgiar-2017-inspire-challenge-da_1.pdf

Chamberlain, J. (2007). Defining smallholder agriculture in Ghana | IFPRI. International Food Policy Research Institute. Retrieved from http://www.ifpri.org/publication/defining-smallholder-agriculture-ghana

Chang, E. (2017, February 14). TaroWorks at Work in the Field with FarmerLink. Retrieved March 23, 2018, from https://taroworks.org/farmerlink-taroworks/

Chau, K. (2017). Cloud Computing for Development: A Four-Stage Analysis of Public, Private, and Hybrid Solutions. Retrieved March 22, 2018, from https://www.ictworks.org/cloud-computing-for-development-a-four-stage-analysis-of-public-private-and-hybrid-solutions/#.Wns3Bq6nHIV

Corea, F. (2017). The convergence of AI and Blockchain: what's the deal? Retrieved March 22, 2018, from https://medium.com/@Francesco_AI/the-convergence-of-ai-and-blockchain-whats-the-deal-60c618e3accc

Data is giving rise to a new economy - Fuel of the future. (2017). Retrieved March 23, 2018, from https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy

Digital Agriculture: Improving Profitability: Accenture Digital Agriculture Service and Connected Crop Solution help the agriculture ecosystem fulfill its potential. (n.d.). Retrieved March 23, 2018, from https://www.accenture.com/us-en/insight-accenture-digital-agriculture-solutions

DititalGlobe Crowdsourcing. (n.d.) Retrieved March 22, 2018, from https://www.digitalglobe.com/products/crowdsourcing

Diskin, P. (1999). Agricultural Productivity Indicators Measurement Guide. The Food and Nutrition Technical Assistance Project (FANTA). USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/Pnacg169.pdf

Employment in agriculture. (2017). Retrieved February 14, 2018, from https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS

Faleide, N. (2017). Will Bitcoin Work in Agriculture? Retrieved March 22, 2018, from http://www.precisionag.com/systems-management/data/will-bitcoin-work-in-agriculture/

Fan, S., Brzeska, J., Keyzer, M., & Halsema, A. (2013). From subsistence to profit: Transforming smallholder farms. In*ternational Food Policy Research Institute.* https://doi.org/10.2499/9780896295582

Fanzo, J. (2017). From big to small: the significance of smallholder farms in the global food system. The Lancet Planetary Health, 1(1), e15–e16. https://doi.org/10.1016/S2542-5196(17)30011-6

FAO. (2016). *Compendium of indicators for nutrition-sensitive agriculture.*

Food and Agriculture Organization (FAO). Agricultural services and digital inclusion in Africa. (2018). FAO. Retrieved from http://www.fao.org/3/a-i7361e.pdf

Fischer, D. (2016). Working with Smallholders: A Handbook for Firms Building Sustainable Supply Chains. IFC. Retrieved from https://www.ifc.org/wps/wcm/connect/8dc5628042112fdbba2fff494779b2ad/Handbook+-+Working+with+Smallholders.pdf?MOD=AJPERES

Fitzgerald, G., & FitzGibbon, M. (2014). A Comparative analysis of traditional and digital data collection methods in social research in LDCs-Case Studies Exploring Implications for Participation, Empowerment, and (mis) Understandings. *IFAC Proceedings Volumes, 47(3*), 11437–11443.

FlyBird Innovations - Irrigation controller. (2014). Retrieved March 22, 2018, from http://www.agricultureinformation.com/postings/flybird-innovations-irrigation-controller/

Gadnis, A. (2017, December). Interview with BanQu.

Gaikwad, G. (2018). What is the difference between cognitive computing and machine learning? Retrieved March 23, 2018, from https://www.quora.com/What-is-the-difference-between-cognitive-computing-and-machine-learning

Gartner Hype Cycle. (2018). Retrieved from: https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

Glover, J. (2017, December). Interview with USAID Office of Agriculture, Research and Policy within Bureau of Food Security.

GSMA. (2016). Mandatory registration of prepaid SIM cards: Addressing challenges through best practice. Retrieved from https://www.gsma.com/publicpolicy/wp-content/uploads/2016/04/Mandatory-SIM-Registration.pdf

GSMA. The Mobile Economy 2017. (2017). Retrieved from https://www.gsmaintelligence.com/research/?file=9e927fd6896724e7b26f33f61db5b9d5&download

Harley, A. (2015). Personas Make Users Memorable for Product Team Members. Retrieved March 23, 2018, from https://www.nngroup.com/articles/persona/

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. https://doi.org/10.1016/j.is.2014.07.006

Hillesland, M. (2017). Guidelines for collecting data for sex-disaggregated and gender-specific indicators in national agricultural surveys. Gender and Development for the Global Strategy to improve Agricultural and Rural Statistics (GSARS). Statistics Division (ESS) of the Food and Agriculture Organization of the United Nations (FAO). Retrieved from http://gsars.org/wp-content/uploads/2017/10/GENDER-FINAL_Guideline_May2017-Completo-10-1.pdf

IBM Brings Watson to Africa for Project Lucy. (n.d.). Retrieved March 23, 2018, from http://www.research.ibm.com/labs/africa/project-lucy.shtml

Informatica. (2013). Preparing for the Big Data Journey: A Strategic Roadmap to Maximizing Your Return from Big Data. Retrieved from https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/white-paper/preparing-for-big-data-journey_white-paper_2342.pdf

Irimia, M. (2016). Five ways agriculture could benefit from artificial intelligence - Watson. Retrieved March 23, 2018, from https://www.ibm.com/blogs/watson/2016/12/five-ways-agriculture-benefit-artificial-intelligence/

IRIS Metrics for Agriculture. (2018). Retrieved from https://iris.thegiin.org/metrics?filters=agriculture%2Cclients

Javaid, U. (2018, January). Interview with Ricult.

Javed, A. (2016). Telenor Pakistan uses Data Science and Analytics to boost mAgri uptake [GSM Association]. Retrieved March 22, 2018, from https://www.gsma.com/mobilefordevelopment/programme/magri/telenor-pakistan-uses-data-science-and-analytics-to-boost-magri-uptake

Jellema, A. (2017). The reference framework for first mile farm data collection, storage and exchange. Retrieved March 22, 2018, from http://www.thefirstmile-referenceframe.info/

Jellema, A. (2018, February). Interview and email exchange with Data-Impact.

Kalita, D., M'Cormack, F., & Heirman, J. (2012). A Literature Review on Farmer Voice. *Institute of Development Studies.* Retrieved from http://pool.fruitycms.com/aline/Downloads/ALINe-Farmer-Voice-Working-Paper.pdf

Karim, L., Anpalagan, A., Nasser, N., & Almhana, J. (2013). Sensor-based M2M Agriculture Monitoring Systems for Developing Countries: State and Challenges. *Network Protocols and Algorithms, 5(3)*, 68–86.

Kelly, C., Manderson, A., & Khoza, T. (2016). Smallholder Farmer Market Segmentation Toolkit: How and why the Toolkit was developed, and lessons learnt from its application in the Mopani and Umkhanyakude districts in South Africa. Southern Africa Food Lab. Retrieved from http://www.southernafricafoodlab.org/wp-content/uploads/2016/11/Case_Study_Final.pdf

Kenya 2017 LR Yield Gap Analysis. (2017). One Acre Fund. Retrieved from https://oneacrefund.org/documents/296/Kenya_2017LR_Yield_Gap_Analysis_v1.pdf

Khamassi, A. (2014). *Building an Analytical Roadmap: A Real Life Example*. Retrieved from https://www.sas.com/content/dam/SAS/en_my/doc/other2/building-analytical-roadmap.pdf

Kilimo Trust. Farmers' Records Book. Regional East African Community Trade in Staples (REACTS) Project. (n.d.). IFAD and Kilimo Trust. Retrieved from https://www.kilimotrust.org/reacts/files/farmer's_record.pdf

Konduru, V. (2017, August 9). The duality of data: What qualitative data can do that Big Data cannot. Retrieved March 23, 2018, from https://towardsdatascience.com/the-duality-of-data-what-qualitative-data-can-do-that-big-data-cannot-3cf3fe52710b

Krishna, P. (2017). What is cognitive analytics? Retrieved March 23, 2018, from https://www.quora.com/What-is-cognitive-analytics

KS, S. (2017, December). Interview with Flybird Innovations.

Laney, D. (2018). Infonomics: *How to monetize, manage, and measure information as an asset for competitive advantage.* Gartner, Inc. Retrieved from https://www.gartner.com/infonomics/

Lee, H. W., & Choudhary, V. (2017, August 17). Agriculture 2.0: how the Internet of Things can revolutionize the farming sector [World Bank]. Retrieved March 22, 2018, from http://blogs.worldbank.org/ic4d/agriculture-20-how-Internet-things-can-revolutionize-farming-sector

Libelium precision farming solution at IBM Watson IoT Center for Cognitive Agriculture application. (2017). Retrieved March 23, 2018, from http://www.libelium.com/libelium-precision-farming-solution-at-ibm-watson-iot-center-for-cognitive-agriculture-application/

Lowder, S. K., Skoet, J., & Raney, T. (2016). The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development*, 87, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041

Lybbert, T. (2018, January). Interview with Mahindi Master.

McKinsey & Company. (2015). Digital America: a tale of have and have-mores. Retrieved from https://www.mckinsey.com/industries/high-tech/our-insights/digital-america-a-tale-of-the-haves-and-have-mores

Manfre, C., Rubin, D., & Nordehn, C. (2017). Assessing How Agricultural Technologies can Change Gender Dynamics and Food Security Outcomes. USAID and INGENAES. Retrieved from http://www.culturalpractice.com/resources/technology-assessment-toolkit/

Manning, L. (2016, September 28). What is the Cybersecurity Threat in Agriculture? Retrieved March 22, 2018, from https://agfundernews.com/what-is-the-cybersecurity-threat-in-agriculture.html

Marita, C., & Karlyn, A. (2018, January). Interview with Mercy Corps Agrifin Accelerate [Skype].

Market-led User-owned ICT4Ag Enabled Information Service (MUIIS). (2017). Retrieved March 23, 2018, from https://g4aw.spaceoffice.nl/en/projects/G4AW-projects/71/market-led-user-owned-ict4ag-enabled-information-service-muiis-.html

Martin, C. (2015). 5 Ways Qualitative Research Can Complement Big Data. Retrieved March 23, 2018, from https://www.flexmr.net/blog/qualitative-research/2015/12/qual-big-data.aspx

Mbo'o-Tchouawou, M., & Colverson, K. E. (2014). Increasing access to agricultural extension and advisory services: How effective are new approaches in reaching women farmers in rural areas? ILRI (aka ILCA and ILRAD). Retrieved from https://cgspace.cgiar.org/bitstream/handle/10568/35495/PR_extension%20services.pdf?sequence=5

Mearian, L. (2018). Blockchain and cryptocurrency may soon underpin cloud storage. Retrieved March 22, 2018, from https://www.computerworld.com/article/3250274/data-storage/blockchain-and-cryptocurrency-may-soon-underpin-cloud-storage.html

Mensink, M., & Vranken, M. (2017). Geodata and ICT Solutions for Inclusive Finance and Food: Innovative Developments --An Overview. Terrafina Microfinance and Fair & Sustainable Advisory Services. Retrieved from http://www.inclusivefinanceplatform.nl/documents/npm-20geodata-20and-20ict-20solutions-20for-20inclusive-20finance-20and-20food-20security.pdf

Merz, C. (2018, January). Interview with the Bill and Melinda Gates Foundation.

Mittal, K., Babcock, L., & Kienzle, L. (Forthcoming). e-Agriculture Promising Practice: Grameen Foundation's FarmerLink. FAO.

Nelson, S., & Swindale, A. (2013). Feed the Future Agricultural Indicators Guide. USAID.

Newman, P. (2017). The Internet of Things 2018 Report: How the IoT is evolving to reach the mainstream with businesses and consumers. Retrieved March 22, 2018, from http://www.businessinsider.com/the-Internet-of-things-2017-report-2017-1

Nordin, K. (2016). Data and analytics for business decision-making: Breakdown of data value chain. Client insights scoping series: Note 2 Breakdown of data value chain: From data to decisions. Insight2impact. Retrieved from http://access.i2ifacility.org/Publications/ScopingNote2_DataValueChain.pdf

Ontologies. (2018). [CGIAR]. Retrieved from http://bigdata.cgiar.org/ontology/

Palmer, T., & Darabian, N. (2017a). A mobile agriculture service by Dialog, Sri Lanka. GSM Association. Retrieved from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/govi-mithuru-mobile-agriculture-service-dialog-sri-lanka.pdf

Palmer, T., & Darabian, N. (2017b). Creating scalable, engaging mobile solutions for agriculture: A study of six content services in the mNutrition Initiative portfolio. GSM Association. Retrieved from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/07/create-scalable-engaging-mobile-solutions-agriculture.pdf

Phillips, D., Waddington, H., & White, H. (2014). Better targeting of farmers as a channel for poverty reduction: a systematic review of Farmer Field Schools targeting. *Development Studies Research, 1(1),* 113–136. https://doi.org/10.1080/21665095.2014.924841

Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences, 114(46),* E9783–E9792.

Principles. (2015). Retrieved from https://opendatacharter.net/principles/

Puranik, M. (2018). Emerging blockchain-based distribution storage market and its effect on cloud computing | Network World. Retrieved March 22, 2018, from https://www.networkworld.com/article/3257706/storage/emerging-blockchain-based-distribution-storage-market-and-its-effect-on-cloud-computing.html

Quisumbing, A. (2014). The results may surprise you: Analyzing sex-disaggregated data. Retrieved March 22, 2018, from http://a4nh.cgiar.org/2014/08/04/the-results-may-surprise-you-analyzing-sex-disaggregated-data/

Raguse, R. (2017). Chatbot with your plants | Agweek. Retrieved March 23, 2018, from http://www.agweek.com/business/agriculture/4335925-chatbot-your-plants

Rakestraw, R. (2016). Precision Ag Innovation Hype Curve. Monsanto Growth Ventures. Retrieved March 23, 2018, from https://drive.google.com/file/d/0B0w-0TiIL5ROMHA5bm1HS3pycVU/view?usp=embed_facebook

Ramasinjatovo, N. (2017, November). ACDI/VOCA Interview [Skype].

Ranger, S. (2018). What is the IoT? Everything you need to know about the Internet of Things right now [ZDNet]. Retrieved March 22, 2018, from http://www.zdnet.com/article/what-is-the-Internet-of-things-everything-you-need-to-know-about-the-iot-right-now/

Representing the worldwide mobile communications industry. (n.d.) Retrieved March 22, 2018, from https://www.gsma.com/

Rosenberg, E. (2017). The Biggest Cloud Computing Companies on Earth. Retrieved March 22, 2018, from https://due.com/blog/biggest-cloud-computing-companies-earth/

Sanghera, A. (2018). Application of Blockchain Technology in Agriculture. Retrieved March 22, 2018, from https://www.linkedin.com/pulse/blockchain-application-agriculture-alekh-sanghera/

Saquib, A. (2018, January). Interview with Viamo Pakistan.

Schmarzo, B. (2016). Developing a big data business strategy. Presentation. Retrieved from https://cdn.oreillystatic.com/en/assets/1/event/144/Developing%20a%20big%20data%20business%20strategy%20Presentation.pdf

Schrage, M. (2016, December 13). Leadership and Big Data Innovation. Retrieved March 23, 2018, from https://hbr.org/webinar/2016/12/leadership-and-big-data-innovation

See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., … Obersteiner, M. (2013). Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. PLoS ONE, 8(7), e69958. https://doi.org/10.1371/journal.pone.0069958

Shekhar, S., Schnable, P., LeBauer, D., Baylis, K., & VanderWaal, K. (2017). Agriculture Big Data (AgBD) Challenges and Opportunities from Farm to Table: A Midwest Big Data Hub Community Whitepaper. Presented at the Midwest Big Data Hub.

Silva, D. (2009). Internet has only just begun, say founders. Retrieved March 23, 2018, from https://phys.org/news/2009-04-internetInternet-begun-founders.html

Smith, O. (2017). Ricult agriculture machine learning. Retrieved from https://techsauce.co/tech-and-biz/ricult-offers-machine-learning-credit-scoring-for-farmers/

Squarcina, M. (2017). Small Family Farms Data Portrait: Basic Information. FAO. Retrieved from http://www.fao.org/fileadmin/user_upload/smallholders_dataportrait/docs/Smallholders_Data_Portrait_-_Methodology_and_Data_Description.pdf

Tatge, J. (2016, July 6). The land grab for farm data. Retrieved March 23, 2018, from http://social.techcrunch.com/2016/07/06/the-land-grab-for-farm-data/

The AIMS Team. (2016). Crop Ontology: harmonizing semantics for phenotyping and agronomy data | Agricultural Information Management Standards (AIMS). Retrieved March 22, 2018, from http://aims.fao.org/activity/blog/crop-ontology-harmonizing-semantics-phenotyping-and-agronomy-data

The Evaluation Design Matrix Template. (n.d.). Bureau for Policy, Planning and Learning. USAID. Retrieved from https://usaidlearninglab.org/sites/default/files/resource/files/evaluation_design_matrix_templates.pdf

The Personal Data Economy. (2018). [Clara Partners]. Retrieved March 23, 2018, from https://www.claropartners.com/project/personal-data-economy/

The Principles. (2018). Retrieved from https://digitalprinciples.org/

The Privacy and Security Principles for Farm Data. (2018). [American Farm Bureau Federation]. Retrieved from http://www.agdatatransparent.com/the-privacy-and-security-principles-for-farm-data/

The Smart Campaign. (2018). Retrieved from https://www.smartcampaign.org/

Thomasson, A., Santos, G., & Basu, A. (2016). Agriculture & Analytics: How would Google farm? Retrieved March 22, 2018, from http://analytics-magazine.org/agriculture-a-analytics-how-would-google-farm/

Tobias, L. (2016). A different approach to crediting smallholder farmers. [FOMIN. IDB.]. Retrieved from https://www.fomin.org/en-us/Home/FOMINblog/Blogs/DetailsBlog/ArtMID/13858/ArticleID/5964/A-different-approach-to-crediting-smallholder-farmers.aspx

Trachtman, R. (2017). Chairperson's Corner: Jump on the PA Bandwagon. Predictive Analytics and Futurism. Retrieved from https://www.soa.org/Library/Newsletters/Predictive-Analytics-and-Futurism/2017/june/2017-predictive-analytics-iss15-trachtman.aspx

Trice, R. (2016). Mapping the AgTech Adoption Curve. Retrieved March 23, 2018, from https://www.forbes.com/sites/themixingbowl/2016/11/02/mapping-the-agtech-adoption-curve/#436e3ed42828

Trucano, M. (2014). Using mobile phones in data collection: Opportunities, issues and challenges [World Bank]. Retrieved March 22, 2018, from http://blogs.worldbank.org/edutech/using-mobile-phones-data-collection-opportunities-issues-and-challenges

Urrea, J. L. (2017, May 17). Big Data is making big impact in agriculture. Retrieved March 22, 2018, from https://ccafs.cgiar.org/blog/big-data-making-big-impact-agriculture

USAID Ghana ADVANCE Wins Digital Development Award for Smartcard Use. (2017). Retrieved March 22, 2018, from http://www.acdivoca.org/2017/10/usaid-ghana-advance-wins-digital-development-award-for-smartcard-use/

Use COAR Controlled Vocabularies to enhance the Interoperability across repositories and with other related systems. (2018). Retrieved March 23, 2018, from http://aims.fao.org/activity/blog/use-coar-controlled-vocabularies-enhance-interoperability-across-repositories-and#at_pco=smlwn-1.0&at_si=5a793ff89942a24f&at_ab=per-2&at_pos=0&at_tot=1

Vaitla, B., Bosco, C., Alegana, V., Bird, T., Pezzulo, C., Hornby, G., … Ruktanonchai, N. (2017). Big Data and the Well-Being of Women and Girls: Applications on the Social Scientific Frontier. Data2x. Retrieved from http://www.data2x.org/wp-content/uploads/2017/03/Big-Data-and-the-Well-Being-of-Women-and-Girls.pdf

Vlamis, T. (2015). The Four Realms of Analytics. Retrieved March 23, 2018, from http://www.vlamis.com/blog/2015/6/4/the-four-realms-of-analytics.html

Wang, T. (2016). Why Big Data Needs Thick Data [Medium]. Retrieved March 23, 2018, from https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7

Waruingi, J., & Muriithi, E. (2016). Digital Harvest. Alliance for a Green Revolution in Africa (AGRA). Retrieved from https://www.raflearning.org/file/1233/download?token=iFv2y-WN

Wayua, C. (2014). IBMVoice: Could Big Data and Cognitive Computing Solve Africa's Greatest Challenges? [Forbes.com]. Retrieved March 23, 2018, from https://www.forbes.com/sites/ibm/2014/11/20/could-big-data-and-cognitive-computing-solve-africas-greatest-challenges/#370194b1227a

What is predictive analytics? (2018). Retrieved from https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/

Wills, A., Pshenichnaya, N., Chorivi, G., Morain, J., & Khada, J. (2015). Micro-Insurance in Mobile Agriculture–Case study & takeaways for the mobile industry. Presented at the GSM Association. Retrieved from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2015/05/ACRE-Africa-Final-Report.pdf

Women, Data & Wisdom: Involving Local Partners in Mobile Data. (2016, June 27). Retrieved March 23, 2018, from https://taroworks.org/involve-partners-mobile-data/

Women's Empowerment in Agriculture Index (WEAI) Resource Center. (2018). Retrieved March 22, 2018, from http://www.ifpri.org/topic/weai-resource-center

World Bank. (2017). I*CT in Agriculture (Updated Edition): Connecting Smallholders to Knowledge, Networks, and Institutions.* The World Bank. https://doi.org/10.1596/978-1-4648-1002-2 https://doi.org/10.1073/pnas.1700319114

World Bank. *Precision Farming: Sensors and Wireless/3G Phone for Smart Irrigation and Fertigation Applications. (2017).* World Bank. Retrieved from https://olc.worldbank.org/content/precision-farming-sensors-and-wireless3g-phone-smart-irrigation-and-fertigation-applications

Young, A., & Verhulst, S. (2017). Aclimate Colombia: Open Data to Improve Agricultural Resiliency. Open Data Impact. Retrieved from http://odimpact.org/files/case-aclimate-colombia.pdf

Yu, Q., Shi, Y., Tang, H., Yang, P., Xie, A., Liu, B., & Wu, W. (2017). eFarm: A Tool for Better Observing Agricultural Land Systems. *Sensors (Basel, Switzerland), 17(3).* https://doi.org/10.3390/s17030453

ANNEXES

Annex 1: Key Informant Interviews

Table 1 below provides a list of participating organization, in alphabetical order, that were interviewed for this assignment.

Table 1: List of institutions interviewed

1	Name	Key Informant
2	ACDI/VOCA	Nirinjaka Ramasinjatovo
3	ADVANCE II	Eric Sunu
4	Amtech Technologies	Pius Sigei
5	Accenture Connected Crop Solutions	Hiran Bhadra, Bhadri Lokanathan
6	AcreAfrica	Muthithi Kinyanjui, Belinda Kaimuri, Lawrence Kinyanjui
7	Akorion	William Luyinda
8	aWhere	Stewart Collis
9	Ayata	Atanu Basu
10	Bankable Frontiers Alliance	David del Sur
11	BanQu	Ashish Gadnis
12	Bill & Melinda Gates Foundation	Christian Merz
13	Better than Cash Alliance	Daniel Waldron
15	CGAP	Jamie Anderson
16	CIAT	Brian King, Daniel Jimenez
17	Farm.ink	Georgia Barrie
18	Federación Nacional de Cafeteros de Colombia (Colombia Coffee Federation)	Juan Pablo Becerra
19	Corpoica	Edwin Rodríguez, Julio Cañas
20	CTA	Benjamin Addom
21	Data-Impact	Andre Jellema
22	ESOKO INSYT	Kwabena Amoateng Frimpong
23	ESOKO TULAA	Gordon Kotei Nikoi
24	Farmerline	Worlali Senyo
25	FinGap-Ghana	Richard Dvorin
26	Flybird Innovations	Satish KS
27	FAO	Henry Burgsteden, Federica Leonarduzzi, Cezar Santos Alvarez
28	Fundación para el Desarrollo Empresarial Rural (FUNDER)	Ángel Meza

	Name	Key Informant
29	Ghana Commercial Agricultural Project	Geoff Anno
30	Gro Intelligence	Hannah Teitelbaum
31	GSMA	Tegan Palmer, Oliver Rowntree
32	ICTworks	Wayan Vota
33	Indian Agriculture Research Institute	Dr. K Rana
34	Indian Society of Agribusiness Professionals (ISAP)	Meghraj Singh
35	Insight2Impact (i2i)	Herman Smit, Mari-Lise du Preez
36	iShamba	Lonita Amisi
37	Mars	Bilal Bawany
38	Mercy Corps AgriFin Accelerate	Collins Marita, Andrew Karlyn
39	Nathan Associates	Ram Tamara
40	NAPL Advisors	Ranjan Sinha
41	One Acre Fund	Anne Maftei, Mike Warmington
42	PHB Associates	Philippe Breul, Lisa Chassin
43	Precision Agriculture for Development	Tomoko Harigaya
44	RAF Learning	Mikael Hook, Clara Collins
45	Ricult	Usman Javaid
46	RiskShield	Agrotosh Mookerjee
47	Sanjhi Tokri	Mayank Gandhi
48	Swisscontact	Ross Jaax
49	TaroWorks	Elaine Chang, Brent Chism
50	Texas A&M University	Alex Thomasson
51	University of California, Davis/Mahindi Master	Travis Lybbert
52	University of Toronto	Ryan Isakson
53	USAID, Office of Agriculture, Research and Policy within Bureau of Food Security	Jerry Glover
54	West Africa Agricultural Productivity Program	Dr. Salakpi von Simeon
55	VIAMO	Abdullah Saqib, Collins Boakye

ANNEX 2: KEY QUESTIONS FOR FARMER PROFILE DATA ASSET MANAGEMENT

McKinsey and Company suggests that three mutually supportive capabilities are necessary to exploit this moment of Big Data (Barton & Court, 2013):

1. Companies must be able to identify, combine, and manage multiple sources of data.
2. Companies need the capability to build advanced-analytics models for predicting and optimizing outcomes.
3. Management must possess the muscle to transform the organization so that the data and models yield better decisions.

Along with McKinsey, several business consulting service providers and technology companies offer resources for developing plans to move into the Big Data economy. (See especially: "Big Data Roadmap: Accelerating Business Returns by Anchoring Your Big Data Journey," 2017; "Preparing for the Big Data Journey: A Strategic Roadmap to Maximizing Your Return from Big Data," 2013; Caire et al., 2017; Khamassi, 2014). The questions listed below have been articulated throughout this document and are summarized here to facilitate reflection.

Defining Smallholders

- How is smallholder farmer defined for your organization? How much flexibility or specificity does your organization require?
- How are smallholder farmers targeted or identified? Would this targeting or identification have implications for how well you reach your targeted beneficiary?
- What definitions or targeting criteria might you consider and what implications might theses have on your strategy?
- What framework/protocol is in place, if any, to capture data about the farm itself? How can your farmer definition map/integrate with other sources of farmer data? (Later sections refer to Ontology and can provide guidance.)

Service Provider Models

- What is your goal? What expectations does your organization have for the use of farmer profile data? Adopting a technology should not be an end in itself. "Too many organizations are making Big Data an IT project instead of making Big Data a strategic business initiative that exploits the power of data and analytics to power the organization's business models." (Schmarzo, 2017)
- What are examples of relevant uses of data within your service provider model that you can draw on for inspiration? Service providers need a "realm of the possible" with respect to integrating new ideas for using farmer profiles (Schmarzo, 2016). Which service providers are using farmer profile data and employing analytics in ways that inspire you? It is hoped that this report will provide a few examples upon which to draw.
- Consider ways in which farmers can interact with the data for greater impact. Blockchain, while not necessarily a data analytic tool, is one such technology promising to put the farmer back in the driver's seat of his or her own data.
- What are the start and end points with regards to funding support and timing?
- Do farmers currently have access to and use their own data in any way? If not, what new technologies might enable and empower farmers to effectively engage with their data?

Type of Data Collected

- When initiating capture of farmer data, service providers must assess which data elements they are prepared and equipped to collect. These should have a natural connection to their organizational goals, and to the purpose of the collection activity (increasing farmer productivity, extending financial services, connecting to purchasers or input providers).

- Service providers should make a concerted effort to understand what data they are already collecting and that is being collected by others and how these can be leveraged and augmented. Seek partnerships where possible.
- Leverage existing ontologies that facilitate cross-organizational data exchanges.
- Envision and articulate a sustainability model. There remains significant uncertainty and flexibility is required. What, if any, opportunities are there to monetize your data? Would others pay for access to your data or its learning products? Are there revenue-sharing models with other partners or stakeholders?

Data Capture Methods

- What methods does your organization currently rely on for capturing farmer data? Do you collect data digitally as an important starting point?
- What technologies could you/should you consider developing farmer profiles?
- Pay attention to the fundamentals of connectivity and usability. These key barriers are often overlooked or minimized. What is the mobile-phone ecosystem where you operate and how might mobile phones be used for data collection?
- New technologies such as satellites, drones, and sensors still face implementation challenges, but they can supplement or replace some data requirements and reduce farmer involvement in data capture.
- What other data sources (such as national information databases) might assist your organization in developing full farmer profiles? What B2B actors might your organization partner with for data sharing?

Cloud storage and computing

- Does your organization currently use cloud storage? Cloud computing? Why not?
- How might your organization benefit from utilizing cloud storage? What data might you access with cloud storage and computing?
- The models for data processing and storage have implications for risk, privacy, and costs. Different models may be needed for pilot and scale.
- What benefits might blockchain technology offer your organization?

Data Analysis

- How does your organization currently analyze its data? Does it use descriptive, diagnostic, predictive, prescriptive, or cognitive analytics?
- How might any of these types of analytics improve your organization's current approach to data analysis? What seems realistic? What is imperative?
- What data science/data analytics skills currently exist within your organization? What gaps in skills exists? What skills should be acquired internally by staff and what outsourced to consultants?
- There is a progression associated with the application of analytics. While it is not necessary to utilize all analytics methods, improved efficiency and applicability can increase relevance for and buy-in of smallholder farmers.

Qualitative data analysis and insights

- Does your organization use Thick Data to develop qualitative archetypes of farmers to understand typical farmers or users of your services? Could you currently do this or would this be a new process?
- How does your organization leverage both qualitative and quantitative data to understand typical and atypical farmer profiles?
- Are there opportunities to better integrate the use of Thick Data with quantitative data being collected and used?

Data Sharing

- What types of data-sharing platforms do you currently use (Google, Box, Dropbox, TaroWorks, etc.)?
- Does your organization have a data-sharing policy? Who is data shared with and how? Is data shared in real time or through periodic reports?
- Is your organization a contributor to or user of others' data? How might you contribute?
- What types of partnerships may be necessary to better leverage others' data? (You do not need to re-create the wheel.)
- What would be the advantages or disadvantages of making your data open?
- What type of data could create more value if it were open source?
- Does your organization have a consumer protection and privacy policy and related processes that ensure policy compliance?
- What risks have you identified for the collection and use of farmer data?
- What misuse of data might be possible that would harm the farmer as well as your organization? How can these be mitigated?

Data Use

- How might farmers be compensated for sharing their data?
- What would you consider your data assets?
- What might other service providers consider as your data assets?
- Data is giving rise to a new economy. Is there an opportunity to monetize your data assets?

Innovation in Farmer Profile Management

- What current combinations of data points, data capture methods, analytics, and data use and management does your organization use? What new combinations might you consider?
- How can you utilize newer data capture methods or analytics to improve your farmer profile data?
- What energy does management have to transform the organization so that the data management practices yield better decisions?
- Start small and lean. With many options to consider, prioritize those that align best with your strategy and provide quick wins and momentum.
- How comfortable is your organization in taking risks? What benefits might there be to being an early adopter of new technologies? What might be the risks?
- For which technologies do you want to see more use cases?
- How might your organization develop a new use case for a data management approach?
- Will utilizing established and known technologies accomplish your goals?